A Neurocognitive Approach to Error Commission and Error Processing in Complex Choice Tasks

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Elisa Caroline Porth

aus Frechen

Abgabe: März 2023

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der

Universität zu Köln im September 2023 angenommen.

Thesis Declaration

We published the first study of this work open access in the journal *Cognitive, Affective, & Behavioral Neuroscience*: Porth, E., Mattes, A. & Stahl, J. (2022). The influence of error detection and error significance on neural and behavioral correlates of error processing in a complex choice task. *Cognitive, Affective, & Behavioral Neuroscience* 22, 1231–1249. https://doi.org/10.3758/s13415-022-01028-6

According to the Contributor Roles Taxonomy (CRediT, Brand et al., 2015), the first author was responsible for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing (original draft), and writing (review and editing). The co-authors contributed to the conceptualization, formal analysis, methodology, software, validation, visualization, writing (review and editing) of this study (André Mattes and Jutta Stahl), as well as to funding acquisition, project administration, resources, and supervision (Jutta Stahl). Minor adjustments to this article were made to match the general layout of the current work.

The second study of this work is currently under revision in Psychophysiology: Porth, E., Mattes, A. & Stahl, J. (2023). *Motor Inhibition versus Interference Suppression: Neural and Behavioural Features of Action Monitoring and Error Processing* [Manuscript submitted for publication]. Department Psychology, University of Cologne.

For this article, the first author was responsible for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing (original draft), and writing (review and editing). The co-authors contributed to conceptualization, methodology, validation, and writing (review and editing) of this study (André Mattes and Jutta Stahl), as well as to funding acquisition, project administration, resources, and supervision (Jutta Stahl).

Abstract

Basic routines and momentous decisions have a constitutive commonality: sooner or later, making a mistake is inevitable. Consequently, a plethora of studies in the domains of psychology and cognitive neuroscience expended great efforts to understand the mechanisms behind error commission more profoundly. The current work comprises two studies that assess the neurocognitive dynamics of error commission and error processing in complex choices while considering error detection, error significance and action inhibition as moderating processes. In our first study, we implemented a novel complex choice task and replicated common variations of neural error processing mechanisms with error detection. We found that an additional self-evaluation rating during half of the experiment enhances neural correlates of error evidence accumulation. Conducting multivariate pattern analyses revealed a broad influence of self-evaluation on whole-brain activity patterns that is potentially associated with attentional resource allocation. In our second study, we assessed the cognitive dynamics of error commission and error processing during complex choices systematically from stimulus onset to post-response adaptation while considering different types of inhibition errors. For motor inhibition we found error-specific variations in neural measures along the entire processing stream, while interference suppression was linked to a particularly resource-intense processing of conflict and inhibition independent of response accuracy. Together, our two studies demonstrate that in complex choices, error-related cognitive processes underlie variations with error detection, self-evaluation and action inhibition that can be uncovered using different behavioural and electrophysiological assessments and a systematic analysing approach. Our findings highlight that the endeavour of investigating error commission and error processing in complex choices using a systematic multimethod approach is pivotal to foster a more profound understanding of error-related cognitive processes.

Table of Content

Introduction

Error commission is an inherent part of everyday life. From action slips over attentional lapses to bad choices with long term consequences, society agrees that things do not always go to plan. An open error culture fosters an adaptive approach, where errors represent an important source of information for improvement. Concealing errors and avoiding a confrontation with mistakes is not in line with this perspective. Instead, thoroughly processing an error is the tenet of an open error culture, as solely the active involvement with a committed mistake allows for future optimization (Wessel, 2018). The ubiquity of error commission and the benefits of deliberate error processing render error-related cognitive processes a recurring focus of psychological research and cognitive neuroscience studies. Over the last decades, research on error processing benefitted from the combination of neuroscientific techniques, multimethodological approaches and eclectic performance parameters (e.g. Allain et al., 2004; Bode & Stahl, 2014; Mattes et al., 2022). As a result, substantial contributions towards a profound understanding of the cognitive mechanisms related to error processing have been made.

A common tool to evoke errors in laboratory settings are speeded choice tasks, where participants are instructed to respond to stimuli by selecting one of two response options. A great variety of studies has assessed error processing from a neurocognitive perspective using such speeded binary response tasks and identified a multitude of factors that are linked to error processing mechanisms in different ways. Importantly, findings suggest that the emergence of error detection is a vital process that is associated with the shape of neural correlates of error processing (Nieuwenhuis et al., 2001; Wessel, 2012). Moreover, results from binary response tasks demonstrated that neural error processing correlates are susceptible to the perceived significance of errors (Hajcak et al., 2005; Riesel et al., 2012). Furthermore, studies discovered that the source of an error has a significant impact on the neural markers of error processing, and that a peculiar vital source of errors are failures in action inhibition (O’Connell et al., 2009). Research identified two types of inhibition errors, i.e. errors due to failures in motor inhibition and errors due to failures in interference suppression (Friedman & Miyake, 2004), that are marked by specific error-related neural features (Riesel et al., 2013).

These brief insights into error-related neurocognitive variations exemplify that binary response tasks are a valuable tool for generating an understanding of basic error-related cognitive processes. However, several studies uncovered that task complexity can have a considerable impact on error processing mechanisms (e.g. Maier et al., 2010; Stahl et al., 2020). While experimental paradigms that capture errors as an outcome of simplified binary choices are elementary for understanding fundamental error-related cognitive processes, the construct of error processing is far more extensive. The emergence of an error in everyday life is usually influenced by a multitude of concurrent alternatives that affect a single decision (Gold & Shadlen, 2007). As of that, it poses a pivotal challenge to investigate errors in the light of more complex choices, as a first step to approximate quotidian decisions. Therefore, in the current work we used a complex choice task to investigate how processing mechanisms vary with error detection, error significance and action inhibition from a neurocognitive perspective.

Theoretical Background

Error Processing and Mental Chronometry

The endeavour to infer from measurable parameters to the temporal dynamics of cognitive processes and to the structure of the cognitive processing system has been frequently described as *mental chronometry* (e.g. Medina et al., 2015). The term has been coined by Posner (1978) in a time where neuroscientific methods were on the raise, but performance parameters were still the main source for inferences about cognitive processes. Mental chronometry has its origins in response time paradigms (Posner, 2005). Response times were regarded as the most suitable measure to gain insights into the chronometry of cognitive processes as they subsume the entirety of cognitive processes from stimulus presentation to response execution. In the early beginnings of mental chronometry research, the division of response times into different stages was used to assess the influence of experimental manipulations more precisely (Donders, 1969) and to narrow down their effects to specific processing stages (Sternberg, 1969). During this period, important questions regarding the architecture of the cognitive processing system were raised that concerned the separability of cognitive processing stages, the production of partial outputs (before process completion), and the nature of the output (discrete or continuous) (Meyer et al., 1988). To this end, these questions still provide content for scientific debates. In the field of error processing research, mental chronometry has been a useful method to elaborate the architecture of the cognitive processing system based on performance parameters.

Behavioural Correlates of Error Commission and Error Processing

The duration of response times serves as an informative indicator for the source of an error (van Driel et al., 2012). When response times for errors are shorter than for correct responses, the error potentially represents a premature response that was executed before thoroughly processing all available information (Ridderinkhof, 2002). In contrast, slower response times for errors than for correct responses indicate an insufficient information quality (Scheffers & Coles, 2000) and lapses of sustained attention (Novikov et al., 2017). Next to response times, further performance parameters such as response force and response corrections deliver additional insights into the nature of an error. More precisely, response force has been associated with inhibition and uncertainty about the correctness of the executed response (Gehring & Knight, 2000; Stahl et al., 2020). Immediate response corrections imply a representation of the response that would have been correct, which in turn indicates that the error probably did not occur because of an insufficient information quality (Maier et al., 2010).

Error commission is not only associated with variations in performance parameters on the trial on which the erroneous response occurs, but also elicits a cascade of processes that carry on to the next trial (Wessel, 2018). In the trial following an erroneous response, we can often observe a slowing in response times, a finding termed post-error slowing (Laming, 1968; Rabbitt, 1968). The phenomenon of post-error slowing was first described as a solely adaptive control mechanism that is aimed at increasing response caution in the next trial (Laming, 1968). The more recent *orienting account* (Wessel, 2018) suggests that error commission elicits an orienting response where the attention is withdrawn from the task at hand and is instead allocated to the identification of the error source. Subsequently, a cascade of adaptive control processes is initiated to increase response caution and to ensure a correct response in the next trial. According to this account, post-error slowing is not solely adaptive as it also captures the emotional occupation with an error and the temporary shift of attention away from the task. Thus, while the response times of an error trial deliver insights into the source of the error, the response times after an error trial deliver information about the extent to which the error was processed.

While mental chronometry research based on response time paradigms advanced, neuroscientific measures gained popularity and became available for a variety of researchers. Concomitant, researchers recognised the potential of conflating electrophysiological methods and mental chronometry (Meyer et al., 1988; Posner, 2005). This combination allowed for unique inferences about the cognitive processing system that were impossible to obtain exclusively through response parameters.

Neural Correlates of Error Processing

The assessment of error-related brain activity allows for inferences about the timing and intensity of associated cognitive processes, which poses a central endeavour of mental chronometry research (Posner, 2005). Recording the electroencephalogram (EEG) is particularly useful to assess neural correlates of error processing, as the high temporal resolution in the range of milliseconds of this non-invasive procedure enables the establishment of real-time onsets of brain activity patterns corresponding to different cognitive processes (Luck, 2014). The EEG is recorded by placing several electrode sites onto the participant's scalp according to the international standard system (i.e. the 10/20 system, Jasper, 1958) and by attaching a reference site to a location with low electrocortical and muscular activity, e.g. onto the mastoid bone behind the ear (Luck, 2014). The scalp electrodes continuously measure voltage differences between the electrocortical signals generated by the brain, which consist of the summation of postsynaptic potentials of cortical pyramidal neurons, and the signal measured at the reference site (Davidson & Jackson, 2000). However, the EEG not only captures brain activity patterns corresponding to the cognitive processes of interest, but also records spontaneous neural activity, muscle activity and ocular movements, which all dilute the electrocortical activity of interest, leading to a low signal-to-noise ratio (Luck, 2014). To increase the signal-to-noise ratio, the electrocortical patterns are time-locked to specific events (e.g. the onset of a response), and are then averaged across several trials of a given task. Through this, unsystematic, task-unrelated variations in the signal which should not coincide with the onset of specific events average out. In the resulting event-related potential (ERP), we can observe various components in the form of positive and negative deflections that research has associated with a variety of cognitive processes (Luck, 2014).

The Error-Related Negativity

The error-related negativity (N_e or ERN) is an ERP component that occurs immediately after the onset of an erroneous response (in the first 150 ms after the response onset, Falkenstein et al., 1991; Gehring et al., 1993). The negative deflection reaches its maximum at frontocentral electrode sites and is thought to be generated in the anterior cingulate cortex (ACC, Dehaene et al., 1994; Gehring et al., 1993). The component is often higher (i.e. more negative) for errors than for correct responses, for which a smaller version of the component can be observed (correct response negativity, CRN or N_c, Falkenstein et al., 1991).

There are several accounts that try to explain the more negative deflection after errors compared to correct responses. The *mismatch theory* constitutes a prominent account that suggests that the N_e reflects the outcome of an internal comparison process between the executed response and the response that would have been correct (Falkenstein et al., 1991; Gehring et al., 1993). For errors, a mismatch between these two responses is detected and signalled by the N_e. An adequate mismatch detection requires a clear representation of the response that was executed and the response that would have been correct. Based on the observations of a weaker but present N_c for correct responses, later amendments to the account suggest that the N_e reflects the internal comparison process itself rather than its outcome (Falkenstein et al., 2000; Vidal et al., 2000). The *conflict monitoring theory* was phrased as an alternative to the mismatch theory and suggests that the N_e signals conflict that is elicited by the simultaneous activation of multiple responses that compete against each other (Botvinick et al., 2001; Yeung et al., 2004). In contrast to the mismatch theory, the conflict monitoring theory does not require a clear representation of the correct response to explain the occurrence of the N_e. In the *reinforcement learning theory*, expectations about performance outcomes evolve through reinforcement learning mechanisms linked to previous responses (Holroyd & Coles, 2002). When performance is worse than anticipated, a respective signal is generated in the basal ganglia. Through a decrease in phasic dopaminergic activity this signal is conveyed to the ACC, where the need for adjustments of motor controllers surfaces as the N_e.

By making assumptions about the timing and architecture of error-related cognitive processes, these theoretical frameworks form the basis for linking electrophysiological findings to the endeavours of mental chronometry research in the field of error processing. Although the accounts place their focus on different processes that are supposed to underlie the N_e, they are not mutually exclusive. In fact, empirical findings can often be interpreted in the light of more than one of these theories without conclusive evidence for the superiority of one account (Gehring et al., 2012). Therefore, elements from all three accounts will be revisited in the course of the current work.

The Error Positivity

The error positivity (P_e) is a positive deflection subsequent to the N_e (approximately 150 to 300 ms after the onset of an erroneous response, Falkenstein et al., 1991; Overbeek et al., 2005). The P_e reaches its maximum at central sites and is discussed to originate in the ACC (Van Veen & Carter, 2002). The *error awareness account* states that the P_e peaks higher for errors than for correct responses (correct positivity, P_c) only when the response is consciously perceived as erroneous (Endrass et al., 2007; Falkenstein et al., 1991; Nieuwenhuis et al., 2001). While the N_e is discussed to reflect fast and automatic processes independent of error awareness (as there is no variation between the N_e for detected and undetected errors, Nieuwenhuis et al., 2001), this account assumes that the P_e captures the subsequent slower process of error awareness based on peripheral response feedback. This assumption is based on the finding that in error detection tasks, the P_e peaks higher for signalled errors (i.e. errors that are detected and signalled by the participant, e.g. with a button press) than for non-signalled errors (i.e. errors that are not reported by the participant). An observed residual positivity after non-signalled errors is interpreted as partial error awareness, where the participant decided against classifying their response as erroneous. The *evidence accumulation account* states that the P_e reflects the process of collecting evidence for an erroneous response (Steinhauser & Yeung, 2010, 2012). The accumulated evidence serves as an input to reach a decision regarding whether or not to signal a response as erroneous, rather than the output of this decision. This account is based on the finding that the P_e varies with the strength of error evidence: The P_e was found to peak higher when the error signalling rate was low and strong evidence was needed to signal an error, and it peaked lower when the error signalling rate was high.

There is an ongoing scientific debate about whether the P_e represents error awareness or error evidence accumulation, or even alternative processes that are not mentioned here for the sake of brevity (Gehring et al., 2012). The current work will focus on the interpretation of the P_e in light of the evidence accumulation account, as this more recent account considers empirical findings that are difficult to explain with the error awareness account (see above, Steinhauser & Yeung, 2010, 2012). Nevertheless, it is important to note that alternative interpretations of the P_e are possible.

Neural Correlates of Error Commission

As outlined previously, a main goal of mental chronometry research is to obtain a holistic picture about the temporal dynamics of cognitive processes and their architecture for the entire decision and response process, from stimulus presentation to the subsequent trials. In this regard, error-related cognitive processes might not start to evolve only with response execution. Instead, the idea of mental chronometry promotes the assessment of ERP components along the entire processing stream. Thus, while the N_e and P_e represent correlates of error *processing* after response execution, the assessment of ERP components before response execution related to visual perception (i.e. the P1 and N1, Haider et al., 1964; Luck et al., 1990; Mangun et al., 1993), conflict and inhibition monitoring (i.e. the N2, Kopp et al., 1996), and attentional resource allocation and response evaluation (i.e. the P3, Polich, 2007) deliver insights into the cognitive architecture of error *commission*.

Moderating Processes: Error Detection, Error Significance and Action Inhibition

Several accounts that aim to explain the cognitive processes behind the N_e and P_e highlight that the neural correlates of error processing are strongly linked to error detection. The mismatch theory assumes that the N_e reflects the emergence of fast and unconscious error detection (Falkenstein et al., 2000; Vidal et al., 2000), and the error evidence accumulation account (Steinhauser & Yeung, 2012) demonstrates that whether or not an error is consciously detected needs to be considered when investigating error-related cognitive processes. As of that, error detection represents a central process that needs to be taken into account when we aim at elaborating our understanding of error-related variations in cognitive processes (for details see Study 1).

To assess whether an error is consciously perceived in experimental tasks, participants are usually instructed to signal whether their response was correct or erroneous immediately after response execution. Importantly, this assessment implies that the participants need to evaluate themselves. This form of self-evaluation requires that the participants actively engage with their own performance and supply a negative evaluation after error commission. In turn, this might enhance the perceived significance of errors for the participants (Grützmann et al., 2014). Findings from binary response task suggest that error significance impacts error processing mechanisms (Maier et al., 2012; Ullsperger & von Cramon, 2004), which makes it an important variable to consider when studying the cognitive dynamics of error processing and error detection (for details see Study 1).

Next to error detection and error significance, the cognitive dynamics of error processing vary as a function of the impeded process(es) that caused the error. For example, when an error arises during a difficult task because the information for response selection is limited (e.g. when stimuli are masked or presented for a short period of time), there can hardly be a clear representation of the correct response to which the erroneous response can be compared to (mismatch theory) or a strong expectation for an accurate performance (reinforcement learning theory). In contrast, when an error occurs during an easy task because a response key was pressed to quickly, but all necessary information for response selection is available and continuously processed, a clear representation of the correct response can emerge (mismatch theory), and the error poses an expectancy violation (reinforcement learning theory). In this case, in line with the theories outlined above, a distinct enhancement of the N_e is expected (Coles et al., 2001). This underlines that the source of an error can have a considerable impact on error processing mechanisms. A prevalently investigated source for errors are failures in action inhibition (O'Connell et al., 2009). Inhibition errors can result from a failure in motor inhibition, where a response is driven by a prepotent but inadequate response tendency, or from a failure in interference suppression, where a response is influenced by irrelevant distractions from the environment (Friedman & Miyake, 2004). Analogously to the precedent example, these two types of inhibition errors display distinct neural features that deserve further investigation (Riesel et al., 2013; for details see Study 2).

Binary and Complex Choice Tasks

In experimental psychology, different tasks are implemented to investigate variations in error processing mechanisms and to establish the underlying cognitive processes. Mental chronometry research initially relied on simple response time paradigms (Posner, 2005). To this end, tasks that are commonly implemented to evoke errors are, amongst others, the Eriksen Flanker Task (Eriksen & Eriksen, 1974), the Go/NoGo task (e.g. O'Connell et al., 2009; van Driel et al., 2012), the Stop-Signal task (Logan, 1994), the Simon task (Simon & Rudell, 1967), the Stroop task (Stroop, 1935), and the Sustained Attention to Response task (Robertson et al., 1997). These tasks elicit errors based on different mechanisms. For example, in a Flanker task, participants are instructed to respond to the central target letter in a string (e.g. HSH or SHS) by pressing one of two response keys (here H and S) while incongruent distractors (i.e. the surrounding letters) induce interference on the level of stimulus processing (e.g. regarding perception and attention). In this task, errors occur when distractor interference is not suppressed sufficiently. In a Go/NoGo task, a response is required for some stimuli, while it needs to be withheld for other stimuli. Here, errors in the form of erroneously executed NoGo responses occur when motor inhibition fails.

Binary response tasks are valuable for inferences on the cognitive processing system as they provide the opportunity to disentangle basic cognitive processes precisely and to target the influence of manipulations on specific cognitive functions. However, they also tend to simplify decisions that, in everyday life, are marked by a considerably higher complexity (Gold & Shadlen, 2007). Thus, especially when the basic cognitive processes behind a construct are already understood relatively well, research needs to evolve and, in a next step, investigate these processes in the frame of more complex tasks.

Advantages of Complex Choice Tasks

Complex choice tasks exceed the structure of binary response tasks by implementing a more complex stimulus-response mapping (Stahl et al., 2020). Combining multiple response options with a larger variety of stimuli increases the number of stimulus-response combinations considerably. This more complex task design has several advantages over a binary response structure. Most importantly, complex choice tasks form a first step towards the aim of relating laboratory research more closely to everyday life. The multitude of response options and the more complex response selection process are (to a certain extent) more relatable to quotidian decisional processes and action selections than binary choices (Gold & Shadlen, 2007). Moreover, in complex choice tasks the spectrum of elicited processes is expanded. For example, while a binary Flanker task captures conflict predominantly on the level of stimulus processing, and a Go/NoGo task captures conflict predominantly on the motor level (Hommel, 2011), a complex choice task has the potential to concurrently subsume both levels of conflict. This increases the variety of processes that can be investigated within a single task, and with that, the similarities to actions in everyday life. Additionally, combining different levels of conflict is an eligible way to provoke a higher number of errors, which is always of interest especially in ERP research to increase the signal-to-noise ratio (Olvet & Hajcak, 2009a). To investigate the cognitive processes underlying error commission and error processing in the context of complex choices, we developed the eight-alternative response task (8ART, Stahl et al., 2020).

The Eight-Alternative Response Task

The 8ART is a speeded choice task with complex stimulus-response assignments (Stahl et al., 2020). The computer-based task consists of eight different stimuli which can appear at eight different locations on the screen and are mapped onto eight different response keys. Complex stimulus-response assignments lead to an equivocal stimulus-response mapping, with a total of 64 possible combinations of stimulus identity and location (for a detailed description of the task see Study 1). Our task roots on the principles of the Simon task, where an incongruence between stimulus identity and response location induces conflict that inevitably increases the probability to commit an error (Simon & Rudell, 1967; Stahl et al., 2020). The 8ART not only induces conflict between stimulus-response assignments, but also between stimuli and, on the motor level, between multiple response options, further increasing the complexity of the task. Due to its complexity, the 8ART appears well-suited to investigate a variety of cognitive processes related to error commission and error processing.

Objectives

The current work investigates the neurocognitive dynamics of error commission and error processing in complex choices. It pursues the central endeavour of mental chronometry research, i.e. to understand cognitive dynamics and structures more profoundly, by combining neuroscientific measures with eclectic performance parameters. In the frame of two studies, we examine how the shape of error-related neural and behavioural parameters vary with error detection, error significance and action inhibition. In both studies we implemented modified versions of the 8ART to investigate these variations in the frame of complex choices. In the first study, we used a multimethod approach by combining behavioural measures and traditional EPR analyses with machine learning techniques to assess how the cognitive dynamics of error processing vary with error detection and error significance in complex choices. The study was published open access in the journal *Cognitive, Affective, & Behavioral Neuroscience* (Porth et al., 2022). In the second study, we used a systematics analysing approach to assess variations in neural and behavioural correlates of error commission and error processing from stimulus onset to post-response adaptation while considering different types of inhibition errors (i.e. motor inhibition errors and interference suppression errors). This study is currently under revision in the journal *Psychophysiology* (Porth et al., 2023).

Study 1

The Influence of Error Detection and Error Significance on Neural and Behavioural Correlates of Error Processing in a Complex Choice Task

Abstract

Error detection and error significance form essential mechanisms that influence error processing and action adaptation. Error detection is often assessed by an immediate self-evaluation of accuracy. Our study used cognitive neuroscience methods to elucidate whether self-evaluation itself influences error processing by increasing error significance in the context of a complex response selection process. In a novel eight-alternative response task, our participants responded to eight symbol stimuli with eight different response keys and a specific stimulus-response assignment. In the first part of the experiment, the participants merely performed the task. In the second part, they also evaluated their response accuracy on each trial. We replicated variations in early and later stages of error processing and action adaptation as a function of error detection. The additional self-evaluation enhanced error processing on later stages, probably reflecting error evidence accumulation, whereas earlier error monitoring processes were not amplified. Implementing multivariate pattern analysis revealed that self-evaluation influenced brain activity patterns preceding and following the response onset, independent of response accuracy. The classifier successfully differentiated between responses from the self- and the no-self-evaluation condition several hundred milliseconds before response onset. Subsequent exploratory analyses indicated that both self-evaluation and the time on task contributed to these differences in brain activity patterns. This suggests that in addition to its effect on error processing, self-evaluation in a complex choice task seems to have an influence on early and general processing mechanisms (e.g. the quality of attention and stimulus encoding), which is amplified by the time on task.

Introduction

Open error culture gives the advice that errors exist to be learned from. We also know that dealing with errors is essential in terms of changing the behaviour towards an intended target state (Wessel, 2018). Unfortunately, we still do not fully understand how errors are processed by the brain. The extent to which error processing and the subsequent adaptation of behaviour take place varies as a function of several variables, with error detection and error significance being key influences. Error detection is usually investigated by an immediate self-evaluation of response accuracy. Based on that, errors can then be divided into errors that the actor signals as such (i.e. signalled errors) and errors that the actor does not signal as such (i.e. non-signalled errors, Maier et al., 2010; Stahl et al., 2020). Several studies have shown that signalled errors do not differ from non-signalled errors in earlier stages of error processing (Nieuwenhuis et al., 2001). However, in later stages of processing, signalled errors seem to be marked by more error evidence accumulation (Steinhauser & Yeung, 2010), which leads to error awareness (Nieuwenhuis et al., 2001).

On a neural level, an earlier stage of error processing is represented by a component of the event-related potential (ERP), the error-related negativity (ERN, Gehring et al., 1993), also termed error negativity (N_e, Falkenstein et al., 1991), which peaks between 0 and 150 ms after an erroneous response and is increased for signalled and non-signalled errors compared with correct responses (correct response negativity/correct negativity [CRN/N_c]) (Falkenstein et al., 1991). A later stage is reflected by the error positivity (P_e, peaking between 150 and 300 ms after an erroneous response) (Falkenstein et al., 1991, Endrass et al., 2007; Overbeek et al., 2005, Steinhauser & Yeung, 2010; Steinhauser & Yeung, 2012), which is increased for signalled errors compared to non-signalled errors and correct responses (Nieuwenhuis et al., 2001). Adaptational control processes often take place after signalled errors, such as in the form of post-error slowing (Laming, 1968; Rabbitt & Rodgers, 1977), while they are less frequently found after non-signalled errors (Klein et al., 2007; Nieuwenhuis et al., 2001; Wessel et al., 2011).

The multifaceted influence of error detection on cognitive processes underlines the importance of a profound understanding of this key variable. To assess error detection, tasks often employ an immediate self-evaluation of response accuracy after each trial. However, this additional task (i.e. the self-evaluation) might modulate error processing itself by enhancing the perceived significance of errors and the attention directed towards them. Previous studies have shown that varying error significance impacts the different stages of error processing, behavioural performance and adaptation mechanisms (Ullsperger & von Cramon, 2004). In a two-choice flanker task, Grützmann et al. (2014) investigated whether the implementation of a response accuracy rating influences error processing. They assumed that if an additional self-evaluation of accuracy enhanced error significance (there error salience) or the attention directed towards errors, this should translate to an increase of N_e amplitude. Alternatively, the authors argued that the attentional resources needed for the additional self-evaluation might lead to a reduction of available resources for the primary task and thus to a decreased N_e amplitude. Grützmann et al. (2014) found an increase in N_e and P_e amplitude when participants were asked to rate their response accuracy, delivering first evidence that self-evaluation might enhance error significance. Other studies delivered similar evidence by showing that early error processing was more pronounced when monetary incentives for correct responses were high, and therefore errors were perceived as more severe, which was reflected by a larger N_e amplitude (Ganushchak & Schiller, 2008; Hajcak et al., 2005; Hsieh et al., 2010; Maruo et al., 2016; Pailing & Segalowitz, 2004; Ullsperger & von Cramon, 2004). Maier and colleagues also supported the error significance account in various studies using the

Eriksen-Flanker-Task (Maier et al., 2008, 2011, 2012; Maier & Steinhauser, 2016). However, other studies did not find evidence for this account (e.g. Maruo et al., 2017; Olvet & Hajcak, 2009b; Paul et al., 2017) and thus challenged the idea of error significance impacting the N_e.

The error evidence accumulation process, reflected by the $P_{e/c}$, might also be affected by error significance in terms of mood modulation (Paul et al., 2017) and monetary punishment (Maruo, 2017; Maruo et al., 2016). Interestingly, the effects of error significance were not necessarily reflected in the participants' performance (*response time*: Ganushchak & Schiller, 2008; Hajcak et al., 2005; Maruo et al., 2016; Paul et al., 2017; *accuracy*: Ganushchak & Schiller, 2008; Hajcak et al., 2005; Paul et al., 2017; Riesel et al., 2012; *post-error slowing*: Paul et al., 2017), although a few studies did report such influences (Grützmann et al., 2014; Hsieh et al., 2010; Maruo et al., 2016; Olvet & Hajcak, 2009b; Riesel et al., 2012). The listed studies overall support the idea that self-evaluation enhances error processing by increasing error significance and the attention towards error processing. However, the majority of the reported studies have implemented response tasks with a binary response mapping (e.g. left-hand and right-hand index finger), which keeps their findings distant from more complex everyday actions. To overcome this limitation, we have developed an eight-alternative response task (8ART, Stahl et al., 2020) that induces a higher cognitive load during response selection. In this task, we provide evidence for variations in neural and behavioural correlates of error processing with error detection given a more complex response selection (Stahl et al., 2020).

In the current study, we used a modified version of the 8ART, which has allowed us to gain further insights into these variations. Using tasks with different characteristics and increasing task complexity are important steps in investigating self-evaluation and error processing to elucidate whether the mechanisms are similar for different task requirements and whether different electrophysiological and behavioural markers might provide evidence for variations in these mechanisms (Olvet & Hajcak, 2009b; Riesel et al., 2015). We added an immediate self-evaluation of response accuracy in the 8ART (Stahl et al., 2020) after each trial as a tool to assess error detection and confidence consecutively. Considering the effects of error significance on error processing and the promising results delivered by Grützmann et al. (2014), the question arises whether or not self-evaluation enhances error processing in a task with a more complex response selection and a higher cognitive load.

Study Objectives

Given that we used a modified version of the 8ART (Stahl et al., 2020) to investigate self-evaluation effects, our first goal was to examine general error detection effects on the neural and behavioural correlates of error processing to gain further insight into error processing in more complex response tasks. Our second goal was to elucidate whether in our more complex task, we were able to replicate that the presence or absence of a self-evaluation in terms of response accuracy modified error processing and participants' performance. While participants simply performed the 8ART in the first part of the experiment (i.e. without self-evaluation), they also had to evaluate their response accuracy on each trial in the second part. If self-evaluation enhances error significance or the attention directed towards errors, then the N_e and P_e amplitudes should be larger for errors that occur when a self-evaluation is required compared to errors after which no self-evaluation is required. If not, then no differences in N_e and P_e amplitudes should be observed between the two conditions. The question of whether or not this effect translates to the participants' performance needs to be explored.

Because univariate ERP analyses did not show stable $N_{e/c}$ effects under previous task settings (only under specific conditions, see Stahl et al., 2020), we further wanted to use a more sensitive method – a machine learning based approach (for details see Methods) – to be able to identify even small differences between the self-evaluation and the no-self-evaluation condition, and their onset if any existed. Based on this approach, Bode and Stahl (2014) have already shown differences between errors and correct responses starting approximately 90 ms before response onset. In the present study, we wanted to investigate whether we were able to obtain similar findings with the 8ART where response selection is much more challenging and whether there were differences between the two evaluation conditions.

Methods

Participants

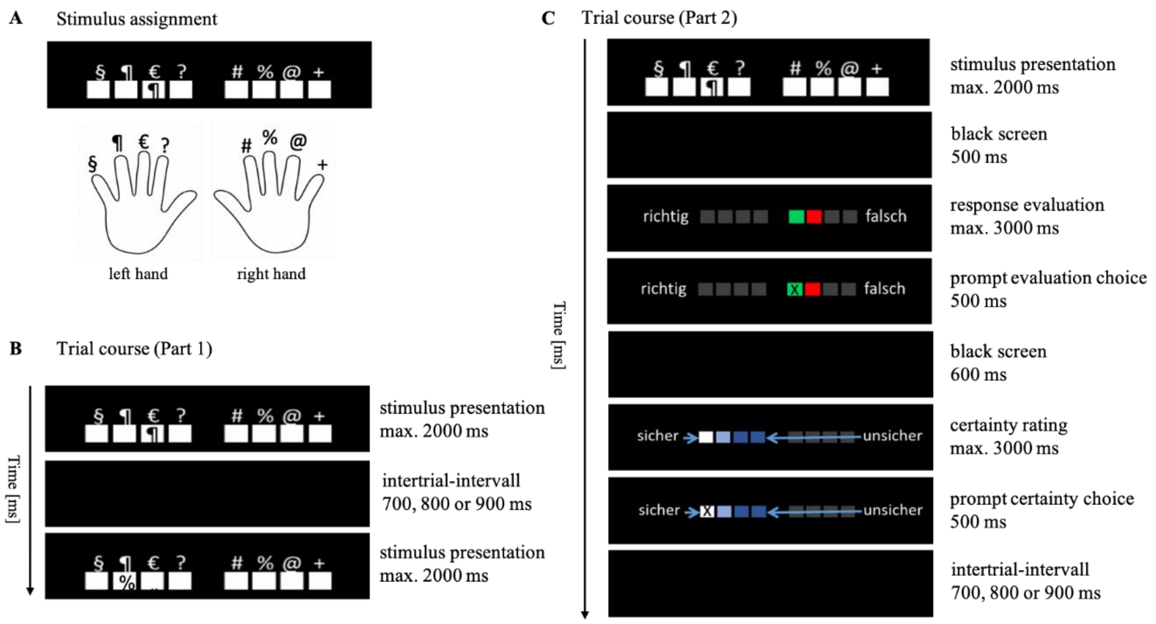
A total of 40 participants, all undergraduate psychology students, took part in the experiment. For the assessment of error processing as a function of error detection (first research goal), the responses were categorized into signalled and non-signalled. Against expectations, only 21 participants (4 males, 17 females, none diverse; age: *M* = 22.4 years, *SD* = 4.5 years) produced enough non-signalled error trials (more than 6, see Olvet & Hajcak, 2009a, for details see Limitations). With this sample size, a Type-I error of 0.05, and a power of 0.80, we were able to uncover effect sizes of $\eta_p^2 \geq 0.14$ (G*Power, Faul et al., 2009). To assess the influence of self-evaluation on error significance (second research goal), it was sufficient to categorize the responses into correct and erroneous. Hence, only seven participants from the entire sample had to be excluded from analyses due to an insufficient number of errors trials (*n* = 6) and an experimenter error (*n* = 1, data saving for EEG was started too late), which led to a final sample size of 33 participants (7 males, 26 females, none diverse; age: *M* = 22.2 years, *SD* = 4.1 years). With this sample size, a Type-I error of 0.05 and a power of 0.80, we were able to uncover effect sizes of $\eta_p \geq 0.11$ (G*Power, Faul et al., 2009). The participants were recruited via a web-based recruitment system (Gotzhein & Elson, 2020), and they received course credit as a reward for participation. All participants reported normal or corrected-to-normal vision. Informed, written consent was obtained from each participant. This study was approved by the ethics committee of the German Psychological Association.

Procedure

The participants completed the 8ART. On each trial, eight white squares appeared on the screen (TFT, 22”) in front of a black background. A different symbol was located above each of the eight squares (from left to right: §, ¶, €, ?, #, %, @, +). In every trial, one of the eight symbols also appeared inside one of the white squares. The participants responded to the symbol inside the white square by pressing one of eight force-sensitive response keys. The keys were positioned in front of the participants. The participants placed their fingers, excluding their thumbs, onto the keys. Each key, and therefore each finger, was assigned to one of the eight symbols above the white squares on the screen. The assignment of the symbols to the participant’s fingers is depicted in Figure 1-1.

Figure 1-1

Stimulus-response assignments and trial course of the eight-alternative response task



Note. (A) Assignment of the symbol stimuli to participant's fingers, (B) the course of one trial from part 1, and (C) the course of one trial from part 2, including response evaluation and certainty rating

The participants were instructed to respond to the symbol appearing in one of the squares with the respective finger that the symbol was assigned to, and they were told to ignore the position of the appearing symbol. They were further instructed to do so as fast and as accurately as possible. If their RT exceeded 1,200 ms, they were given the feedback "even faster" (German: *noch schneller*), which appeared in red font in the middle of the screen before the next trial started. The participants did not receive any feedback on the accuracy of their response. The time course of one trial is illustrated in Figure 1-1 (B: no-self-evaluation condition, C: self-evaluation condition).

Stahl et al. (2020) measured error detection and detection certainty on an 8-point rating scale ranging from "certainly right" to "certainly wrong" and found that participants tended to mainly use the extreme values of the scale. Therefore, we split the original scale into two parts and assessed error detection and response certainty separately. First, the participants evaluated whether their response was correct or erroneous by pressing the force-sensitive key with their index or middle finger of the right hand, respectively. For this evaluation, the participants were given a response time limit of 3,000 ms. Next, the participants were asked to rate their certainty of their former evaluation using the four keys assigned to their left hand. The rating ranged from certain to uncertain and was supported by a presentation of matching squares on the screen. For this rating, the participants were again given a response time limit of 3,000 ms. We did not counterbalance the responding hands for the two-staged rating procedure to avoid confusing the participants (as reported by some participants after the experiment of Stahl et al., 2020). We ensured that the inter-trial interval between the primary task and the first rating was long enough for error processing mechanisms to proceed before any possible shifts in topography due to the build-up of readiness potentials could occur and that the time window for the rating response was long enough to prevent overlaps with processes related to previous or subsequent responses. After the certainty rating, the trial was completed, and the next trial started. The inter-trial interval randomly varied in length (700, 800, or 900 ms). The type and position of the symbols appearing in the white squares were randomized and distributed equally throughout each block of the experiment.

The experiment consisted of two parts. During the first part, there was neither a self-evaluation nor a certainty rating but only the main task (i.e. *no-self-evaluation condition*). We did not prolong the inter-trial interval in this condition so that it matches the length of the evaluation and certainty rating as such a long black screen interval evokes mind wandering, detachment from the task, boredom, and fatigue. These processes would limit the comparability of the two self-evaluation conditions. The aggregated cognitive load during the main task in the self-evaluation condition is probably higher than in the no-self-evaluation condition due to the anticipation that one has to evaluate one’s own performance, which is the process we wanted to manipulate. We did not implement an additional task in the no-self-evaluation condition as compensation since an additional task would set off other cognitive processes we cannot control. The no-self-evaluation condition consisted of seven blocks with a total of 394 trials (approximately 15 minutes, including breaks). The first two blocks were shorter practicing blocks with no RT limit; the other 5 blocks consisted of 64 trials each. Within a block, each symbol-position combination occurred once. The evaluation and the certainty rating were displayed in the second part of the experiment (*self-evaluation condition*). This condition consisted of 11 blocks with 64 trials each, leading to a total number of 704 trials (approximately 60 minutes, including breaks). Between blocks, the participants were given the opportunity to rest and proceed with the next block whenever they felt ready. The second part was longer than the first part, because a higher number of trials was needed to obtain a sufficient number of signalled and non-signalled responses. The order of the two parts was not counterbalanced because if the self-evaluation part had preceded the no-self-evaluation condition, participants would have potentially kept evaluating their responses – even if the specific instruction had been omitted.

Apparatus

Force sensors (FCC221-0010-L, DigiKey MSP6948-ND) that were embedded in each of the eight keys registered the response that the participants executed. The response signal of the sensors was digitized by a VarioLab AD converter at a sampling rate of 1,024 Hz. The program also captured the real-time onset of the stimuli on the computer screen via a photo sensor attached to the screen, which reacted to a change of brightness. The keypress was registered (also in real time) as a response if it exceeded a threshold of 40 cN. The response keys were adjusted to the individual length of the participant’s fingers, and they were calibrated according to the individual weight of the fingers. To prevent an excess of head movement and to ensure a constant distance to the screen (86 cm), the participant’s chin rested on a chin support during the experiment.

Electrophysiological Data

The EEG was recorded by 63 active, unipolar Ag/AgCl electrodes (Acticap, Brain Products) that were placed onto the scalp according to the standard international 10-20 system (Jasper, 1958) (FP1, FP2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C3‘, C1, Cz, C2, C4, C4‘, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, PO9, O1, Oz, O2, PO10). The electrodes were online referenced against an electrode placed on the left mastoid and they were re-referenced offline against the mean of the signal recorded at the left and right mastoid. Passive, bipolar Ag/AgCl electrodes (ExG-Amplifier, Brain Products) were used to record the electrooculogram (EOG). They were positioned horizontally next to the outer side of both eyes near the temples (horizontal EOG), as well as vertically above and below the left eye (vertical EOG). The EEG and EOG signals were continuously recorded by a BrainAmp DC amplifier (Brain Products) at a sampling rate of 500 Hz with a filter from DC to 70 Hz.

For the pre-processing of the ERP data, we locked the intervals of each trial onto the response onset with a time window ranging from 100 ms before and up to 800 ms after the response onset. We performed a baseline correction based on the 100 ms before the response onset (for MVPA 100 ms before stimulus onset). We implemented an ocular correction algorithm to correct for eye blinks in the activity patterns (Gratton et al., 1983). With an artefact rejection, we removed trials in which the ERP waves exceeded a threshold of ± 100 µV (*M* = 10.8%, *SD* = 12.2% of all trials that lay within the RT limit). We performed a current source density (CSD) analysis to detach the signal from the reference electrodes and to reduce the redundant activity of neighbouring electrodes (Perrin et al., 1989). We did not apply any high-pass or low-pass filters to the data to avoid temporal smearing and amplitude reductions. The $N_{e/c}$ amplitude was determined at the FCz electrode as the mean amplitude ±2 data points around the negative peak (i.e. we averaged the activity of five data points: the data point of the peak and two data points before and after that) in a time window of 0 ms to 150 ms after response onset. The $P_{e/c}$ amplitude was determined at the Cz electrode as the mean amplitude ± 2 data points around the positive peak and mean activity in a time window of 150 ms to 300 ms after response onset. Topographical maps (see *Results*) confirmed these electrodes as the local extrema of the two components.

Multivariate Pattern Analysis

In addition to the traditional univariate analyses of ERP components at individual electrode sites, the ERP signal can also be analysed in a multivariate fashion (i.e. at all electrode sites). Multivariate pattern analysis (MVPA) enables certain time windows to be identified, at which a person’s response can be reliably predicted from their ERP data (Bode et al., 2012). Including information from all electrode sites enables a more exhaustive use of information regarding the participant’s brain activity patterns, which in turn allows for the identification of even small differences between the self-evaluation and the no-self-evaluation condition in the current study.

The MVPA were performed by using the support vector machine classification (SVC) from the Decision Decoding Toolbox (Bode et al., 2019). First, all analyses were performed for each participant separately (first-level analyses). Second, these results were averaged across all participants and inference statistics were computed (second-level analyses). We conducted a series of analyses, each contrasting two conditions: (1) no-self-evaluation condition: correct responses vs. errors; (2) self-evaluation condition: correct responses vs. errors; (3) correct responses: no-self-evaluation condition vs. self-evaluation condition; and (4) errors: no-self-evaluation condition vs. self-evaluation condition. The data were time-locked to the response onsets (i.e. the response to the symbol stimuli) with a time window ranging from 900 ms before response onset to 200 ms after response onset, which should cover the entire process from stimulus presentation to the overt response. In the first-level analysis, these time windows were subdivided into smaller non-overlapping time windows with a length of 10 ms, each including five data points. The data points within these smaller time windows and of all electrodes were then transformed into vectors that formed the spatio-temporal pattern of the two conditions. A linear classifier was then trained with these vectors using the LIBSVM toolbox (Chang & Lin, 2011). The SVC process was submitted to a tenfold cross-validation, resulting in a total of 100 classification processes. The classification accuracies of these runs (i.e. the percentage of cases in which the classifier was able to sort the vectors into the conditions they belonged to) were then averaged to obtain one classification accuracy value for each of the 10-ms time windows of each participant in the respective comparison condition. For each participant, in addition to the classification accuracy, the chance classification was computed based on the empirical data. This resembled the previous procedure, except that the classification probability was obtained by permutations within a near-identical shuffled-labels analysis (Bode et al., 2019).

The second-level analyses were conducted separately for each 10-ms time window, for which the classification accuracies were averaged across all participants. The chance classification was also averaged across all participants. The mean classification accuracy and the chance classification were then tested against each other for every 10-ms time window within 900 ms before to 200 ms after the response onset with a series of Bonferroni-corrected *t*-tests. If the mean classification accuracy deviated significantly from the chance level, then the classifier was able to decode the condition from the brain activity patterns in the respective time window.

Statistical Analyses

For the investigation of error detection in the self-evaluation condition, we conducted univariate ANOVAs with repeated measures for the factor Response Type (signalled correct, signalled error, and non-signalled error) for response rates, the percentage of multiple responses (where response force exceeded 40 cN on more than one key), RT (time span between the stimulus onset and the exceedance of 40 cN on one of the keys), response force (RF; max. force in a trial), post-response correct responses (percentage of trials followed by a correct response), evaluation RT (time span between the evaluation screen onset and the exceedance of 40 cN on one of the keys), response certainty, $N_{e/c}$ and $P_{e/c}$ mean amplitudes, and $P_{e/c}$ mean activity. We applied Greenhouse-Geisser correction when the sphericity assumption was violated and Tukey's HSD for within-comparisons as post-hoc tests. We compared signalled and non-signalled errors regarding error correction (percentage of errors that were immediately followed by the correct response) with a two-tailed *t*-test. The pre-post response time difference (pre-post RT_{diff}) was computed by subtracting the RT following an erroneous trial from the RT preceding an erroneous trial and then calculating the mean of these differences. This procedure was described as a robust measurement for post-error slowing (Dutilh et al., 2012). For the sake of comparison, the pre-post RT_{diff} also was calculated for correct responses. We used three Bonferroni-corrected one-tailed *t*-tests to contrast pre-post RT_{diff} for signalled errors and correct responses, for non-signalled errors and correct responses, and for signalled errors and non-signalled errors. Similar to many studies (see Wessel, 2012 for discussion), we had to exclude the non-signalled correct responses from our analyses, because the majority of the participants showed less than six trials in this condition and a reliable quantification of the $N_{e/c}$ could not be ensured (Olvet & Hajcak, 2009a). To assess the influence of self-evaluation, we compared the behavioural and neural parameters of the no-self-evaluation condition with the self-evaluation condition. For this, we performed a series of univariate ANOVAs with repeated measures for the factors Self-Evaluation (no self-evaluation, self-evaluation) and Accuracy (correct, error) for RT, RF, pre-post RT_{diff}, post-response correct responses, $N_{e/c}$ and $P_{e/c}$ mean amplitudes, and $P_{e/c}$ mean activity.

Results

Our first goal was to investigate error detection effects on the behavioural and neural correlates of error processing. Therefore, in a first step we report results of the self-evaluation condition, where we categorized responses into the three response types *signalled correct*, *signalled errors*, and *non-signalled errors*. Our second goal was to examine the effect of self-evaluation on participants' performance and on error processing. Thus, in a second step we report results of the contrast between the no-self-evaluation condition and the self-evaluation condition.

Response Types

The self-evaluation condition was designed to assess neural and behavioural correlates of error processing as a variation of error detection. The descriptive statistics for the behavioural and neural parameters split by the three response types are depicted in Table 1-1.

Table 1-1

Means and standard error of means for all assessed behavioural and electrophysiological variables separately for each response type (signalled correct responses, signalled errors and non-signalled errors) for n = 21 for the second part of the experiment

	Signalled Correct		**Signalled Errors**		**Non-signalled Errors**	
	M	**SE**	**M**	**SE**	**M**	**SE**
Response rates[a] [%]	88.6	1.8	7.4	1.2	4.0	0.7
Multiple responses [%]	4.9	1.2	12.1	1.9	44.8	4.5
Response time [ms]	814.7	14.1	835.2	22.6	820.2	31.2
Response force [cN]	162.8	10.2	135.4	7.6	88.3	5.7
Pre-post RT_{diff} [ms]	0.1	2.8	32.8	12.3	85.0	28.5
Post-response correct [%]	69.2	3.7	69.4	4.5	65.9	5.8
Evaluation RT [ms]	391.5	29.4	656.8	50.8	489.2	35.4
Certainty rating [1-4][b]	3.7	0.1	3.5	0.1	2.9	0.1
$N_{e/c}$ amplitude [µV/cm²]	-0.16	0.03	-0.26	0.04	-0.23	0.04
$P_{e/c}$ amplitude [µV/cm²]	0.13	0.03	0.33	0.05	0.22	0.03
$P_{e/c}$ mean activity [µV/cm²]	0.11	0.03	0.26	0.04	0.15	0.03

[a] Relative to responses from the remaining three response types. Responses which exceeded the response time limit of 1,200 ms were excluded (11.4 ± 1.7% of all trials). Among response types, only 1.0 ± 0.3% were non-signalled correct responses.

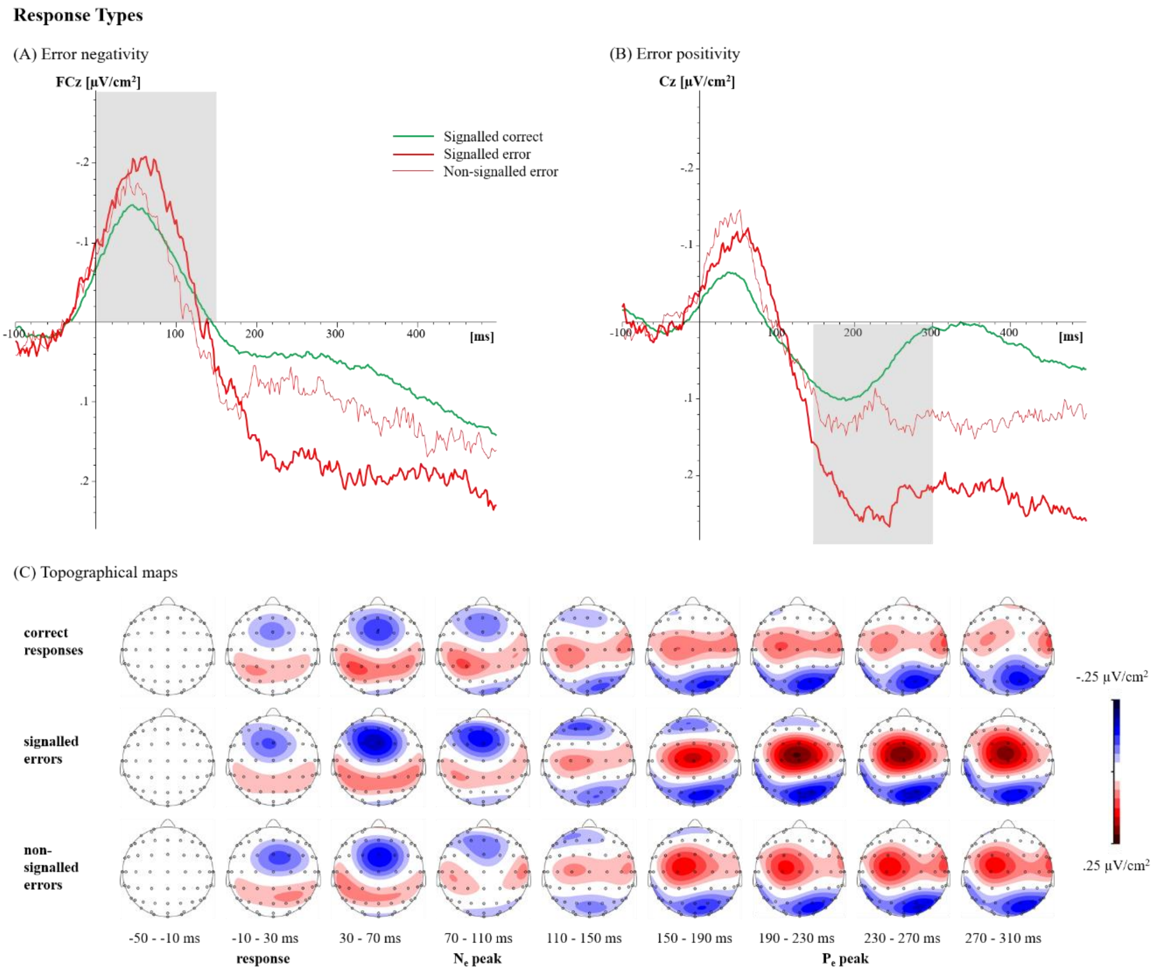
[b] Certainty values ranged from 1 = very uncertain to 4 = very certain

Behavioural Results. Response rates differed significantly between response types, $F(1.10, 21.94) = 928.46$, $p < .001$, $\varepsilon = .55$, $\eta_p^2 = .98$. From the three response types, signalled correct responses occurred more frequently than signalled and non-signalled errors, both p values < .001, whereas the two error types did not differ significantly, $p = .278$. The percentage of multiple responses varied significantly with response type, $F(1.27, 25.40) = 80.83$, $p < .001$, $\varepsilon = .63$, $\eta_p^2 = .80$. There were significantly more multiple responses for non-signalled errors compared to the other two response types, both p values < .001, whereas the signalled errors and signalled correct responses did not differ significantly from each other, $p = .094$. In non-signalled error trials a correct response followed more often (43.6 ± 4.6%) than in signalled error trials (9.5 ± 1.6%), $t(20) = -8.97$, $p < .001$, $d = -1.96$. The RTs did not differ significantly between response types, $F(1.45, 29.09) = 0.33$, $p = .653$, $\varepsilon = .73$, $\eta_p^2 = .02$. The peak RF significantly varied by response type, $F(2, 40) = 41.59$, $p < .001$, $\varepsilon = .83$, $\eta_p^2 = .68$, with signalled correct responses being the most forceful, followed by signalled errors and non-signalled errors, all $p < .005$. The pre-post RT_{diff} was significantly larger for signalled errors compared to correct responses, $t(20) = -2.37$, $p = .042$, $d = -0.52$, and larger for non-signalled errors compared to correct responses, $t(20) = -3.02$, $p = .010$, $d = -0.66$. The signalled errors and non-signalled errors did not differ significantly in pre-post RT_{diff}, $t(20) = -1.64$, $p = .348$, $d = -0.36$. The percentage of post-response correct responses did not vary significantly with response type, $F(2, 40) = 1.06$, $p = .355$, $\varepsilon = .80$, $\eta_p^2 = .05$. The Evaluation RT varied significantly between response types, $F(2, 40) = 27.63$, $p < .001$, $\varepsilon = .85$, $\eta_p^2 = .58$. Evaluation RT was smallest for signalled correct responses, followed by non-signalled errors and signalled errors, all p values < .026. The ANOVA for response certainty yielded a significant effect for response type, $F(1.23, 24.51) = 22.02$, $p < .001$, $\varepsilon = .61$, $\eta_p^2 = .52$, indicating less certainty for non-signalled errors compared to signalled errors and signalled correct responses, both p values < .001. Signalled errors did not differ significantly from signalled correct responses, $p = .134$.

Electrophysiological Results. The $N_{e/c}$ mean amplitude differed significantly between response types, $F(2, 40) = 11.76$, $p < .001$, $\eta_p^2 = .37$, with signalled errors and non-signalled errors having a larger amplitude than signalled correct responses, $p < .001$ and $p = .005$, respectively. The $N_{e/c}$ amplitude of signalled and non-signalled errors did not differ significantly, $p = .377$. The $P_{e/c}$ mean amplitude differed significantly between response types, $F(2, 40) = 18.11$, $p < .001$, $\eta_p^2 = .48$, with signalled errors and non-signalled errors having a larger amplitude than signalled correct responses, $p < .001$, and *p* = .027, respectively. Post-hoc tests also revealed a larger amplitude for signalled errors than for non-signalled errors, *p* = .005. The $P_{e/c}$ mean activity differed significantly between response types, *F*(2, 40) = 13.02, $p < .001$, $\eta_p^2 = .39$, with signalled errors having a higher activity than non-signalled errors, *p* = .003, and signalled correct responses, $p < .001$, which did not differ significantly from each other, *p* = .360. The CSD-transformed averaged ERP waveforms together with their topographical maps are depicted in Figure 1-2 (for untransformed ERP data see Supplement).

Figure 1-2

Averaged CSD-transformed ERP waveforms for the three response types



Note. Averaged CSD-transformed waveforms of the ERP components (A) error negativity, measured at FCz electrode as the mean amplitude ± 2 data points around the negative peak in the time window of 0-150 ms after response onset and (B) error positivity, measured at Cz electrode as the mean amplitude ± 2 data points around the positive peak and as mean activity in the time window of 150-300 ms after response onset separately for each response type (signalled correct, signalled and non-signalled errors), as well as the respective topographical maps

No Self-evaluation vs. Self-evaluation

In the second set of analyses, we compared the self-evaluation condition with the no-self-evaluation condition (Table 1-2).

Table 1-2

Means and standard error of means for all assessed behavioural and electrophysiological variables separately for each condition (no self-evaluation vs self-evaluation) for correct and erroneous responses for n = 33

	No Self-evaluation				**Self-evaluation**			
	Correct		**Error**		**Correct**		**Error**	
	M	**SE**	**M**	**SE**	**M**	**SE**	**M**	**SE**
Response rates [a] [%]	87.1	1.3	12.9	1.3	88.9	1.6	11.1	1.6
Response time [ms]	787.9	14.0	813.4	16.7	796.1	12.4	804.4	16.5
Response force [cN]	146.2	8.2	104.6	3.8	156.5	8.3	107.0	3.7
Pre-post RT_{diff} [ms]	-9.4	2.2	52.5	9.1	-0.7	1.6	34.5	8.9
Post-response correct [%]	77.1	2.0	75.2	3.1	78.3	2.5	76.5	2.9
$N_{e/c}$ amplitude [µV/cm²]	-0.12	0.02	-0.20	0.03	-0.13	0.02	-0.16	0.02
$P_{e/c}$ amplitude [µV/cm²]	0.08	0.02	0.19	0.03	0.07	0.02	0.21	0.03
$P_{e/c}$ mean activity [µV/cm²]	0.02	0.02	0.10	0.02	0.02	0.02	0.15	0.02

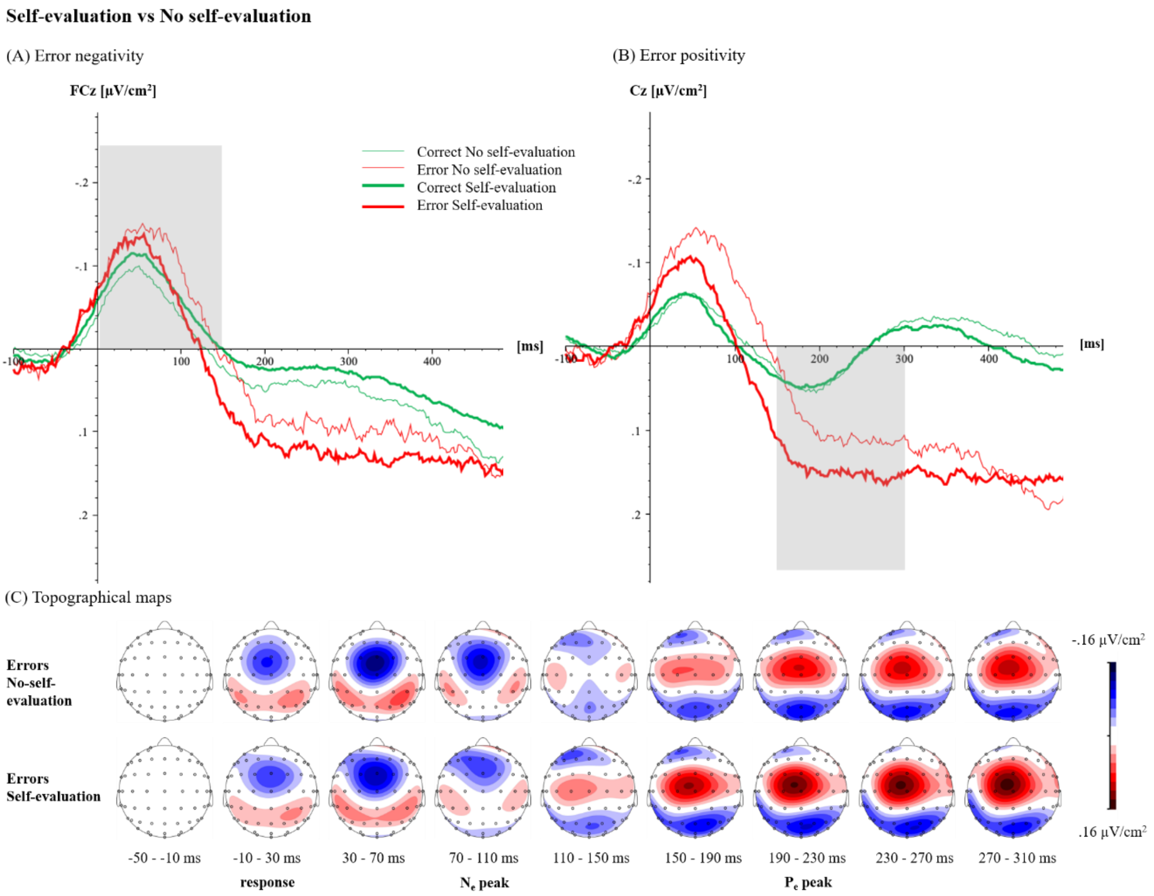
[a] Relative to all responses within the response time limit; responses which exceeded the response time limit of 1,200 ms were excluded. The percentage of too slow responses did not differ between the two evaluation conditions (no-self-evaluation condition: 10.6% ± 1.6, self-evaluation condition: 9.9% ± 1.4, $t(32) = 1.14$, $p = .264$, $d = 0.20$)

Behavioural Results. For RT, there was no significant main effect of self-evaluation, $F(1, 32) = 0.00$, $p = .960$, $\eta_p^2 < .01$, no significant main effect of accuracy, $F(1, 32) = 1.98$, $p = .169$, $\eta_p^2 = .06$, and no significant interaction, $F(1, 32) = 1.56$, $p = .221$, $\eta_p^2 = .05$. The ANOVA for peak RF did not yield a significant main effect of self-evaluation, $F(1, 32) = 1.98$, $p = .169$, $\eta_p^2 = .06$, but a significant main effect for accuracy, $F(1, 32) = 58.13$, $p < .001$, $\eta_p^2 = .64$, and an effect for the self-evaluation by accuracy interaction, $F(1, 32) = 5.38$, $p = .027$, $\eta_p^2 = .14$. Post-hoc tests showed that correct responses were more forceful than errors in the no-self-evaluation condition, $p < .001$, and in the self-evaluation condition, $p < .001$. Correct responses did not differ significantly between self-evaluation conditions, $p = .159$. Errors also did not differ significantly between self-evaluation conditions, $p = .961$. For the pre-post RT_{diff} of errors, there was no significant difference between the no-self-evaluation and the self-evaluation condition, $t(32) = 1.23$, $p = .229$, $d = 0.21$. For the post-response accuracy there was no significant main effect of self-evaluation, $F(1, 32) = 0.79$, $p = .382$, $\eta_p^2 = .02$, no significant main effect of accuracy, $F(1, 32) = 2.14$, $p = .153$, $\eta_p^2 = .06$, and no significant interaction, $F(1, 32) = 1.01$, $p = .940$, $\eta_p^2 < .01$.

Univariate Electrophysiological Results. For the $N_{e/c}$ mean amplitude, there was no significant main effect of self-evaluation, $F(1, 32) = 1.33$, $p = .257$, $\eta_p^2 = .04$, but there was a significant main effect of accuracy, $F(1, 32) = 16.72$, $p < .001$, $\eta_p^2 = .34$, with the amplitude being higher for errors than for correct responses. A significant interaction between self-evaluation and accuracy, $F(1, 32) = 5.85$, $p = .021$, $\eta_p^2 = .15$, revealed a tendency for a larger N_e amplitude in the no-self-evaluation condition compared to the self-evaluation condition, $p = .075$, whereas there was no significant difference in the N_c between the self-evaluation conditions, $p = .881$. Furthermore, only in the no-self-evaluation condition was the N_e amplitude significantly larger than the N_c amplitude, $p < .001$, whereas they did not differ significantly in the self-evaluation condition, $p = .119$. For the $P_{e/c}$ mean amplitude, there was no significant effect for self-evaluation, $F(1, 32) = 0.47$, $p = .500$, $\eta_p^2 = .01$, but there was an effect for accuracy, $F(1, 32) = 31.23$, $p < .001$, $\eta_p^2 = .49$, with the amplitude being higher for errors than for correct responses. The interaction between self-evaluation and accuracy was not significant, $F(1, 32) = 1.65$, $p = .208$, $\eta_p^2 = .05$. For the $P_{e/c}$ mean activity, there was no significant effect for self-evaluation, $F(1, 32) = 3.68$, $p = .064$, $\eta_p^2 = .10$, but there was an effect for accuracy, $F(1, 32) = 26.70$, $p < .001$, $\eta_p^2 = .45$. The interaction between self-evaluation and accuracy was significant, $F(1, 32) = 6.27$, $p = .018$, $\eta_p^2 = .16$. Post-hoc tests revealed a higher mean activity for errors than for correct responses in the no-self-evaluation condition, $p = .007$, and in the self-evaluation condition, $p < .001$. The $P_{e/c}$ mean activity also was higher for errors in the self-evaluation condition than in the no-self-evaluation condition, $p = .017$, whereas correct responses did not differ significantly between conditions, $p > .999$. The CSD- transformed averaged ERP waveforms together with their topographical maps are depicted in Figure 1-3. We report results for the untransformed ERP data in the Supplement.

Figure 1-3

Averaged CSD-transformed ERP waveforms for the two self-evaluation conditions



Note. Averaged CSD-transformed waveforms of the ERP component (A) error negativity, measured at FCz electrode as the mean amplitude (±2 data points) around the negative peak in the time window of 0-150 ms after response onset and (B) error positivity, measured at Cz electrode as the mean amplitude (± 2 data points) around the positive peak and as mean activity in the time window of 150-300 ms after response onset grouped by accuracy (correct, error) and experimental condition (no self-evaluation, self-evaluation), as well as the respective topographical maps for errors of both conditions

Multivariate Electrophysiological Results. Comparing errors and correct responses within the no-self-evaluation condition (Figure 1-4 A) and within the self-evaluation condition (Figure 1-4 B) did not reveal any time windows at which the classifier was able to significantly predict the response accuracy (errors vs. correct) from the brain activity patterns (for all 10-ms time windows, the predictions of the response accuracy were around chance level). However, the algorithm was successful in predicting the evaluation conditions (self-evaluation vs. no self-evaluation) from brain activity within correct trials (Figure 1-4 C) and within error trials (Figure 1-4 D) significantly above chance-level. Interestingly, the classifier was able to predict the evaluation conditions already starting at 550 ms and 650 ms before response onset (within correct trials and within error trials, respectively). The respective feature weights analyses are displayed in the Supplement. In sum, this machine-learning based approach, which identifies time windows where condition-specific brain activity is decodable from the spatio-temporal activity patterns (i.e. ERP data points from all electrodes within 10-ms time windows), showed that the self-evaluation and the no-self-evaluation condition differ in these patterns at a very early stage of processing and during the entire period even after the response.

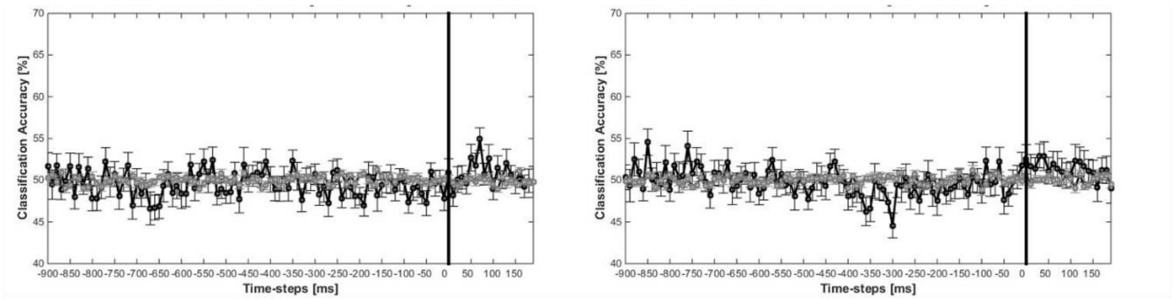
Figure 1-4

Response-locked MVPA results

Classification Accuracies

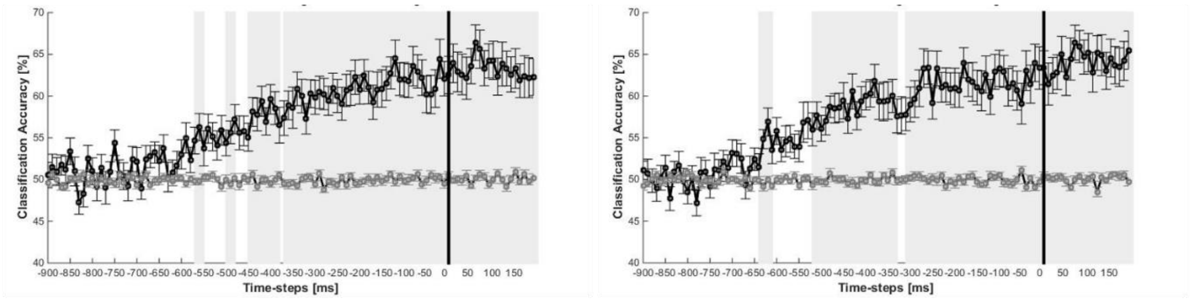
(A) No self-evaluation: Correct vs Errors

(B) Self-evaluation: Correct vs Errors



(C) Correct: No self-evaluation vs Self-evaluation

(D) Errors: No self-evaluation vs Self-evaluation



Note. Classification accuracies in percent for each time step for comparisons between correct responses and errors in the no-self-evaluation condition (A) and the self-evaluation condition (B), as well as for comparisons between evaluation conditions for correct responses (C) and errors (D). Grey areas indicate time windows in which the classification accuracies (black lines) differ significantly from the empirical chance level (grey lines). Corrected for multiple comparisons.

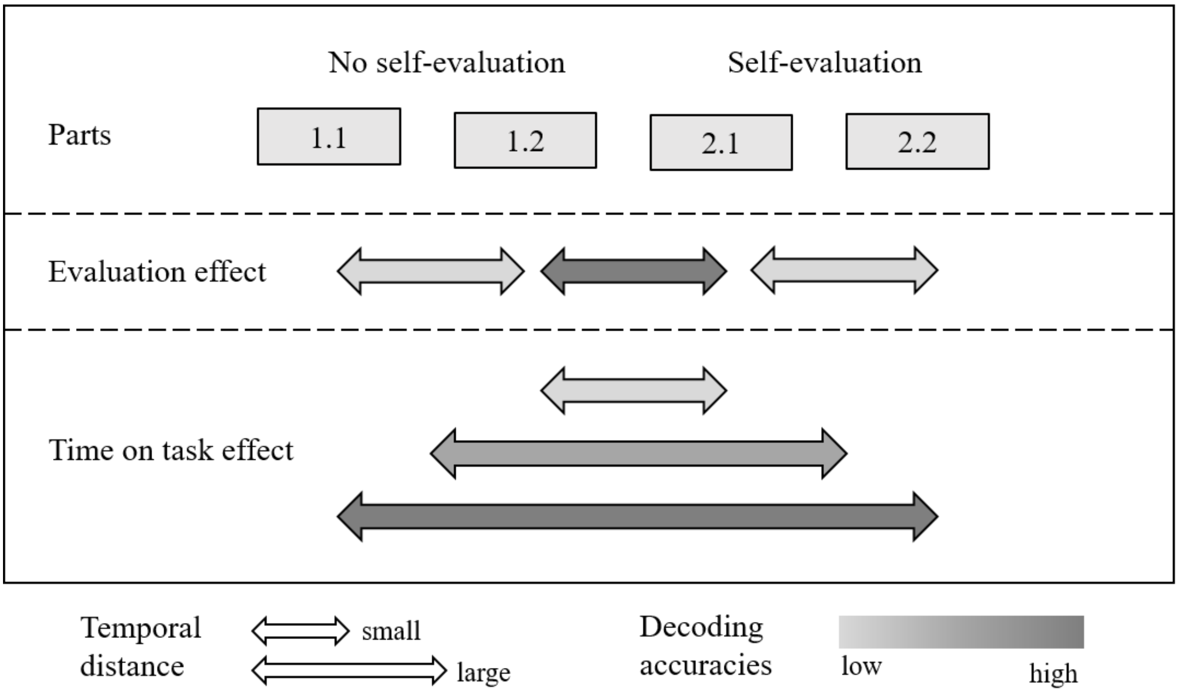
The additionally performed MVPA including RT-matched comparisons, accounting for differences in RTs between conditions, reflected similar patterns with significant time windows starting 500 ms before response onset for correct responses and 600 ms before response onset for errors (which is not shown for the sake of brevity).

Subsequent Exploratory Analyses

The remarkable differences in brain activity patterns several hundred milliseconds before response onset between the self-evaluation conditions, irrespective of the response type, could indicate a broader influence of self-evaluation on brain activity patterns preceding a response. However, due to the constant order of both conditions, we cannot rule out the possibility that the differences reflect changes in more general processing mechanisms over the course of the experiment (e.g. training, attention, vigilance). To further assess whether the MVPA results reflect an effect of self-evaluation, an effect of time on task or a combination of both, we carried out several subsequent exploratory analyses. We made a within-condition split, resulting in four parts (1.1, 1.2, 2.1 and 2.2), which we then contrasted against each other using MVPA (for an overview, see Figure 1-5). If the previous MVPA results are based on a time on task effect, the classification accuracies for contrasts between two parts in our subsequent MVPA should increase with temporal distance between the two contrasted parts (i.e. part 1.1 vs. part 2.2 should show the highest classification accuracies). If the MVPA effects were due to self-evaluation, the classification accuracies should be higher for contrasts between cross-conditional parts compared to contrasts between within-conditional parts, which have the same temporal distance. In other words, accuracies should be higher for contrasting part 1.2 with part 2.1 than for contrasting the parts within the no-self-evaluation condition (1.1 vs. 1.2) and contrasting the parts within the self-evaluation condition (2.1. vs. 2.2). Accordingly, increased decoding accuracies for temporally more distant parts and increased decoding accuracies for the cross-conditional parts would suggest and influence of both time on task and self-evaluation.

Figure 1-5

Implication of a self-evaluation effect and a time-on-task effect



Note. Assumptions regarding the classification accuracies of the contrasted parts based on the split-logic of both conditions. Evaluation effect: Classification accuracies are higher for cross-condition comparisons compared to within-condition comparisons of the same temporal distance. Time on task effect: Classification accuracies are higher for contrasts between parts with large temporal distance compared to parts with small temporal distance.

Because the differences in brain activity patterns were not error specific, we collapsed errors and correct responses for the subsequent exploratory analyses. This prevents an exclusion of a substantial number of datasets due to an insufficient number of error trials. It also enables the inclusion of the 33 datasets that we analysed previously. We also analysed behavioural measures for the four parts. We conducted additional analyses of behavioural data in a blockwise manner (5 blocks for the no-self-evaluation condition and 11 blocks for the self-evaluation condition) to test for time on task effects in the participants' behaviour. We analysed RT and RF collapsed over response types. To capture time on task effects on the level of response rates, we also analysed the error rates of the different parts and blocks.

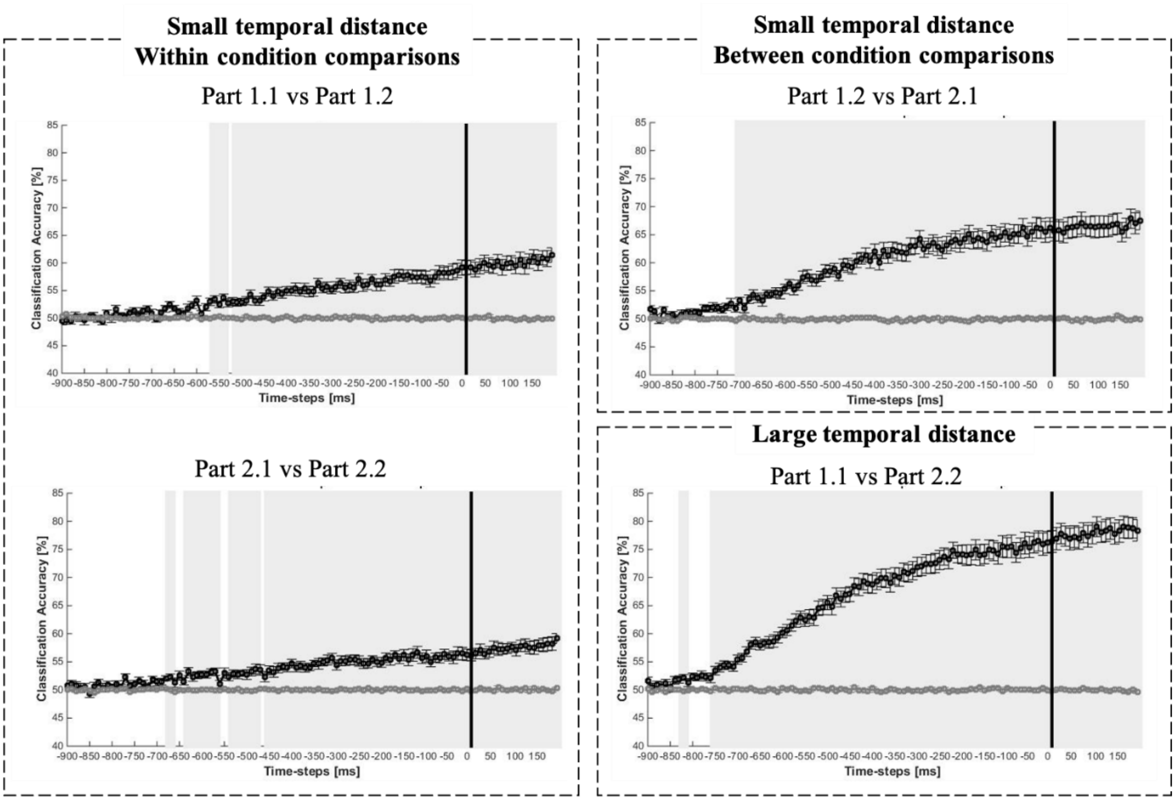
Behavioural Results. A one-way ANOVA with the factor part (1.1, 1.2, 2.1, 2.2) showed no significant variation for error rate, $F(1.95, 62.40) = 1.87$, $p = .163$, $\varepsilon = .65$, $\eta_p^2 = .06$. RT varied significantly by part, $F(1.83, 58.65) = 4.92$, $p = .013$, $\varepsilon = .61$, $\eta_p^2 = .13$. Post-hoc tests revealed that RT was slower in part 1.1 (804.9 ± 13.2 ms) than in part 1.2 (779.6 ± 14.4 ms), $p = .003$, and slower in part 1.2 compared with part 2.2 (799.0 ± 12.3 ms), $p = .032$. The other parts did not differ significantly, all p values > .201. The peak RF did not differ significantly between parts, $F(1.60, 51.17) = 2.06$, $p = .147$, $\varepsilon = .53$, $\eta_p^2 = .06$. Comparing the blocks within the two self-evaluation conditions, neither the no-self-evaluation condition, $F(3.15, 100.88) = 1.77$, $p = .156$, $\varepsilon = .79$, $\eta_p^2 = .05$, nor the self-evaluation condition, $F(5.45, 141.75) = 0.98$, $p = .439$, $\varepsilon = .55$, $\eta_p^2 = .04$, showed a significant variation in error rates with block. The ANOVA for RT in the no-self-evaluation condition yielded a significant effect of block, $F(4, 128) = 6.15$, $p < .001$, $\varepsilon = .83$, $\eta_p^2 = .16$. Post-hoc tests revealed that RTs were slower for the first block compared with the last two blocks, both p values < .005, and RTs in the second and the third block were both slower than in the last block, both p values < .032. The other blocks did not differ significantly, all p values > .071. In the self-evaluation condition, RT did not differ significantly between blocks, $F(6.39, 204.35) = 0.86$, $p = .530$, $\varepsilon = .64$, $\eta_p^2 = .03$. In the no-self-evaluation condition, there was a significant effect of block on peak RF, $F(2.82, 90.26) = 3.78$, $p = .015$, $\varepsilon = .71$, $\eta_p^2 = .11$. Post-hoc tests revealed that responses were more forceful in the first block compared with the last two blocks, both p values < .028. The other blocks did not differ significantly in peak RF, all p values > .175. In the self-evaluation condition, for the peak RF there was no significant effect of Block, $F(4.16, 133.12) = 1.43$, $p = .225$, $\eta_p^2 = .04$.

MVPA Results. The MVPA results are depicted in Figure 1-6. To test whether the classification accuracies are significantly higher for certain decoding groups (i.e. contrasts between two parts) compared with others, we computed aggregated classification accuracy scores for each individual, decoding group, and condition. To this end, we subtracted the empirical chance accuracy from the classification accuracy for each time step of each decoding group for each participant. The resulting difference scores indicate the percentage of additional trials that the classifier successfully decoded correctly when given the non-shuffled compared with the shuffled data. Finally, we averaged these scores across time steps. For every participant, each decoding group is now represented by a single difference value that represents its mean classification accuracy under consideration of chance accuracy. With these difference scores, we conducted a one-way ANOVA with repeated measures for the factor decoding group to test whether the classifier achieved higher classification accuracies for certain decoding groups compared with others.

The ANOVA showed that the aggregated accuracy scores indeed varied significantly between decoding groups, $F(2.87, 91.72) = 96.11$, $p < .001$, $\varepsilon = .57$, $\eta_p^2 = .75$. Post-hoc tests revealed that the accuracy scores were higher for the cross-condition decoding group with small temporal distance (part 1.2 vs. 2.1, $M = 10.2\%$, $SE = 1.1\%$; i.e. the classifier sorted on average an additional 10.2% of the trials into the correct condition compared to chance classification) compared with both within-condition decoding groups with small temporal distances (part 1.1 vs. 1.2, 5.0 ± 0.6%; and part 2.1 vs. part 2.2, 4.2 ± 0.5%), both *p*s < .001, whereas the within-condition decoding groups did not differ, $p = .907$. Post-hoc tests also revealed that the scores were highest for the decoding group contrasting the temporally most distant parts (part 1.1 vs. part 2.2, 17.4 ± 1.2%), followed by the decoding groups contrasting parts of medium temporal distance (part 1.1 vs. 2.1, 14.4 ± 1.2% and part 1.2 vs. part 2.2, 14.1 ± 1.1) and then the decoding groups comparing parts of the smallest temporal distances, all *p* values < .003. The two decoding groups that compare parts of medium temporal distance did not differ, *p* = .999. The respective feature weights analyses are displayed in the Supplement.

Figure 1-6

MVPA results for part-wise comparisons



Note. Classification accuracies in percent for each time step for comparisons between the parts with small temporal distance within conditions (part 1.1 vs. part 1.2 and part 2.1 vs. part 2.2), with small temporal distance between conditions (part 1.2 vs. part 2.1) and with the largest temporal distance (part 1.1 vs. part 2.2). Grey areas indicate time windows in which the classification accuracies (black lines) differ significantly from the empirical chance level (grey lines).

Discussion

Using the recently developed 8ART (Stahl et al., 2020), we examined the neural and behavioural features of error processing in a complex task with challenging response selection. In the first step, we assessed variations in neural and behavioural correlates of error processing with error detection in this novel task. The second step investigated the influence of self-evaluation in terms of response accuracy on error processing by extending traditional univariate ERP analyses with the more modern MVPA approach. Finally, we explored how the behavioural parameters and MVPA results varied with time on task.

Error Detection and Error Processing in the 8ART

In our modified version of the 8ART, we were able to uncover variations in error processing mechanisms with error detection. This supports the previous findings in the literature from more simple task designs. In contrast to Stahl et al. (2020), we found the common $N_{e/c}$ effect (Nieuwenhuis et al., 2001) of response type in our study irrespective of response speed, whereas Stahl et al. (2020) only found an $N_{e/c}$ difference for fast trials. Due to the structure of the present study, we suppose that the representation of the stimulus-response mapping was much more trained by the first part, which comprised more than 300 trials (Stahl et al. only used 64 practice trials). Therefore, a better-established memory representation might have enabled a higher sensitivity of error processing mechanisms related to the $N_{e/c}$. The $P_{e/c}$ effect was replicated in terms of an increased P_e amplitude for signalled errors compared to non-signalled errors and correct responses. This supports the notion that error evidence accumulation proceeds similarly for binary (Steinhauser & Yeung, 2010; Steinhauser & Yeung, 2012) and more complex choice tasks (Stahl et al., 2020). The behavioural results fit the findings of Stahl et al. (2020), which indicates that our task modifications did not affect the previously reported patterns.

Interestingly, the non-signalled errors in our task stand out from the common literature findings regarding P_e amplitude, evaluation RTs, and pre-post RT_{diff} (Endrass et al., 2007, 2012; Klein et al., 2007; Nieuwenhuis et al., 2001; Wessel et al., 2011), which were all larger than for correct responses. The high number of multiple responses and response corrections in non-signalled errors appear to be responsible. These additional responses might have induced uncertainty about response accuracy (e.g. when participants pressed the correct key immediately after an incorrect key) and impeded error evidence accumulation reflected by a smaller but not missing P_e amplitude, and (ultimately) prevented successful error signalling. Indeed, non-signalled errors were associated with the lowest response certainty. In our complex paradigm, error evidence accumulation is probably time consuming particularly for non-signalled errors due to the high number of multiple responses. This was reflected in the evaluation RTs (signalled errors > non-signalled errors > correct responses), which suggests that the time spent on response evaluation is an indicator for the amount of error evidence that is accumulated over time. Thus, the gradual variation of the $P_{e/c}$ (reflected by $P_{e/c}$ amplitude) might be more realistic compared with an all-or-nothing principle (reflected by $P_{e/c}$ mean activity). Endrass et al. (2012) also found evidence for the $P_{e/c}$ being a gradual measure of error awareness. Yet, an enhancement only for signalled errors is commonly reported in literature (Nieuwenhuis et al., 2001), which underlines the impact of different ERP quantifications and the importance to discuss deviations between them.

Finally, non-signalled errors feature a large pre-post RT_{diff}, i.e. post-error slowing, which might have served as a mechanism to reduce the uncertainty associated with non-signalled errors in the next trial. This contrasts previous findings of increased post-error slowing for signalled but not for non-signalled errors (Endrass et al., 2007, 2012; Klein et al., 2007), even when more response corrections for non-signalled errors were observed (Endrass et al., 2007). In our study, the consecutive evaluation process (detection and certainty ratings) might have given participants the opportunity to process their response more deeply and to continue error processing after an initial detection response. Therefore, although error evidence might not have been sufficient by the time the response was inaccurately signalled as correct, subsequent error processing might have led to error detection and therefore increased post-error slowing in the next trial. Post-error slowing then reflected the prolongation due to orienting and adaptational mechanisms elicited not only by the previous incorrect trial but also by the previous incorrect response signalling, which would support a relation of post-error slowing to the perceived correctness of a response (Endrass et al., 2012).

Self-evaluation and Error Processing

Our second goal was to investigate whether self-evaluation influences error processing and behavioural performance. Comparing the self-evaluation and the no-self-evaluation condition did not reveal differences regarding behavioural measures. Assuming that self-evaluation increases error significance, this is in line with the finding that the effects of error significance do not always translate to the participants' performance (Ganushchak & Schiller, 2008; Hajcak et al., 2005; Maruo et al., 2016). Similarly, Grützmann et al. (2014) did not find significant differences in error rates and post-response accuracy improvement between the two self-evaluation conditions. However, while post-error slowing was showing in the no-self-evaluation condition, the researchers reported slower RTs in general and no post-error slowing in the self-evaluation condition. They concluded that a more cautious response mode in the self-evaluation condition led to a ceiling effect in RTs that hindered post-error slowing. In our study, the RTs in the two evaluation conditions were similar and therefore the absence of RT ceiling effects may explain the observed post-error slowing in both the self-evaluation and the no-self-evaluation condition. Interestingly, the pattern of evaluation RTs we found in our study (signalled error > non-signalled error > signalled correct) fits the interpretation of Grützmann et al. (2014), who also observed slower evaluation RTs for errors than for correct responses. This pattern indicates that post-error slowing already occurs during the self-evaluation process and reflects an orienting response to expectancy violating events (Grützmann et al., 2014; Wessel, 2018).

Surprisingly, we found an $N_{e/c}$ difference only in the no-self-evaluation condition and a tendency for a reduced N_e in the self-evaluation condition. In case of a clear effect of self-evaluation induced by an enhanced error significance or attention to error monitoring, error processing should have been reinforced in the self-evaluation condition. The inversed effect, which is opposed to what Grützmann et al. (2014) found, can have various reasons. First, the literature has shown that error monitoring is modulated by task difficulty (Kaczkurkin, 2013; Riesel et al., 2015; Schreiber et al., 2012; Somon et al., 2019; Voodla & Uusberg, 2021) and uncertainty (Bultena et al., 2020; Pailing & Segalowitz, 2004; Scheffers & Coles, 2000; Selimbeyoglu et al., 2012). While the N_e amplitude is found to be attenuated, the N_c amplitude increases with task difficulty and uncertainty, presumably due to a hindered error detection process (Pailing & Segalowitz, 2004). Our results showed that the difference between the N_e and N_c in the self-evaluation condition was significantly smaller than in the no-self-evaluation condition and that the N_e tended to be smaller in the self-evaluation condition than in the no-self-evaluation condition. These findings suggest that an effect of task difficulty or uncertainty might have counteracted an effect of error significance. However, studies that reported $N_{e/c}$ modulations also reported modulations of behavioural parameters by task difficulty (Pailing & Segalowitz, 2004; Riesel et al., 2015; Schreiber et al., 2012; Somon et al., 2019; Voodla & Uusberg, 2021)). An effect of task difficulty was not reflected in our behavioural data, because the two self-evaluation conditions did neither differ significantly regarding error rates, RTs, post-response slowing, and post-response accuracy improvement nor regarding the percentage of too slow responses, which speaks against an increased task difficulty in the self-evaluation condition. Moreover, our participants reported an overall high certainty in their responses, and there was no enhancement of N_c amplitude, which also argues against a higher task difficulty in the self-evaluation condition.

Second, we cannot preclude that the decrease in N_e amplitude might indicate an effect of time on task (Boksem et al., 2006; Horváth et al., 2020; Tops et al., 2006; Tops & Boksem, 2010). However, the part- and blockwise analyses of the behavioural data revealed neither systematic increases in error rates nor systematic variations in RT (RT increase as an indicator of fatigue; continuous RT decrease as an indicator of practice), which we would expect in case of a relevant time on task effect (Horváth et al., 2020; Tops & Boksem, 2010). In addition to that, it has been shown that increasing the motivation to engage with the task, e.g. by implementing external monetary incentives, counteracts a time on task induced N_e decrease (Tops & Boksem, 2010). We assume that our self-evaluation rating increased participants' motivation to direct more attention to the task to prevent a negative self-evaluation, reducing the likelihood of a mere time on task effect as the single cause for a decrease in N_e amplitude and the absence of a significant $N_{e/c}$ difference in the self-evaluation condition.

Third, individual differences also can account for the absence of an error significance effect. Previously, it has been shown that for some individuals, errors have a higher significance and are processed more deeply (Riesel et al., 2012). For example, punishment substantially modified error monitoring in highly anxious participants (Riesel et al., 2012, 2019). In this notion, it is possible that in our more complex task, an effect of error significance might only unfold in individuals that display an increased sensitivity to modulations of error significance. Hence, including traits such as anxiety (Riesel et al., 2012), perfectionism (Stahl et al., 2015), or narcissism (Mück et al., 2020) as moderators in future studies will help to further elucidate the relationship between self-evaluation and error processing in our more complex task.

Last and most importantly, we used two scales: the binary detection scale and the four-point certainty evaluation. The consecutive and thus prolonged self-evaluation process in our task might have absorbed more attentional resources than the single three-level rating scale Grützmann et al. (2014) used in their study. This might have led to less cognitive resources available for error monitoring in the primary task and therefore to a decreased N_e amplitude in the self-evaluation condition. Grützmann et al. (2014) argued that, although they found an increased N_e amplitude in their self-evaluation condition indicating an enhanced error significance, it is still possible that this effect was weakened by the absorption of attentional resources through the rating procedure, but the influence of error significance was more impactful. In our task where the ratings were more complex, it might in fact have been the other way around.

We did not find a variation of $P_{e/c}$ amplitude with self-evaluation. However, $P_{e/c}$ mean activity showed a clear effect of self-evaluation, with a higher mean activity for errors in the self-evaluation condition compared with errors in the no-self-evaluation condition. This indicates that self-evaluation increases the need for processing errors more deeply and the motivation to intensify the evidence accumulation process. While early error processing mechanisms reflected by the $N_{e/c}$ might have been reduced because the subsequent rating procedure absorbed attentional resources, error evidence accumulation reflected by the $P_{e/c}$ might be exactly the process that absorbs the attentional resources. We assume that error evidence accumulation is the relevant process to successfully detect the response accuracy. Therefore, more attentional resources might have been directed towards error evidence accumulation, while they were withdrawn from early, fast error monitoring.

Self-evaluation and Whole-brain Activity

Interestingly, the more sensitive classifier was unable to identify a difference in whole-brain activity pattern between errors and correct responses. Instead, it identified differences between the two self-evaluation conditions before and after the response. The difference between the findings of Bode and Stahl (2014), who reported significant error-related differences in a two-alternative task, might be related to the more complex nature of our task. Due to the high number of response alternatives and multiple responses, brain activity patterns of correct responses and errors might have been more similar to each other and intermixed with other processes. The complex response selection and the more complex stimulus-response assignment might have induced more “noise” and thus challenged the decoding of an equivocal error-related pattern. Surprisingly, the self-evaluation conditions seemed to differ already on early processing stages (more than 500 ms before response onset). Because we already precluded error-specific differences as a cause for this effect, we conducted additional MVPA to investigate the factors that contributed to the remarkable differences. Subsequent analyses revealed that the large differences between the two evaluation conditions were in parts due to time on task, because the classifier was more successful in identifying the different experimental parts when the temporal distance between them was larger. Fluctuations in processes, such as fatigue, vigilance, and attention, might be the underlying sources for this effect. Moreover, self-evaluation seemed to amplify this established time on task effect, because the classifier also was more successful in differentiating self-evaluation from no-self-evaluation parts than parts from the same condition. One possible explanation for this might be that the two self-evaluation conditions differ in the allocation of attentional resources. While the primary task was the same for both conditions, there might have been less attentional resources available for the primary task in the self-evaluation condition due to the complex rating procedure. This might explain the early differences between the two evaluation conditions and is in line with our ERP results. We can preclude the influence of confounds such as error rates, RT, and RF on our MVPA results, because these did not vary systematically with the observed MVPA patterns.

Limitations

Despite our successful detection of behavioural and neural differences dependent on response type, as well as our interesting findings regarding self-evaluation, our study is marked by some limitations. First, the second part of the experiment had more trials than the first part, which enables a sufficient number of signalled and non-signalled error trials but limits comparability. We decided against prolonging the first part to avoid fatiguing the participants. To ensure that the measurement of the $N_{e/c}$ is equally stable in both conditions, we conducted trial-matched ERP analyses, which we report in the Supplement.

Second, unfortunately yet unsurprisingly (for a discussion see Wessel, 2012), our participants did not produce a sufficient number of non-signalled correct responses, which led us to exclude this response type from analyses.

Third, we had to exclude a considerable number of participants from the analyses, because they did not produce enough non-signalled errors, whereas a larger sample size would have been preferable. We further inspected the data sets to ensure the effectiveness of our task design and data quality for the first set of analyses. In the past, a low number of non-signalled errors has been a common issue in error awareness studies (Kirschner et al., 2021) that often led to the exclusion of a substantial number of participants (Niessen et al., 2020; O'Connell et al., 2007; Scheffers & Coles, 2000). In a recent study, we obtained a reasonable rate of non-signalled errors with our newly developed 8ART (Stahl et al., 2020, 5.6 ± 0.7% non-signalled errors, 5 out of 38 participants excluded due to an insufficient number of non-signalled errors) and thus showed that the task is well-suited to investigate error awareness. To preclude systematic differences between the data sets included in our study and those excluded due to an insufficient rate of non-signalled errors, we compared both subsamples. The subsamples did not differ significantly in behavioural parameters regarding RT, peak RF, or post-response slowing (for statistics see Supplement). Instead, the first part of our study probably led to a more mature error-signalling process in the second part, which might be responsible for the rate of low non-signalled errors. Therefore, either further increasing task complexity to foster non-signalled errors or carefully reducing the opportunity for practice will be an important step in our future studies.

Fourth, the blockwise analyses show a change in performance that is stronger in the beginning of the experiment. This shift in performance is limited to RT, whereas the error rates are stable across the experiment. This nonlinear change in performance most likely reflects an effect of practice. This might be a problem for the partwise MVPA that imply a more linear change of neural activity with time on task. To ensure that the partwise MVPA are not distorted by the nonlinear change in RT, we conducted RT-matched partwise MVPA. The results are reported in the Supplement. The analyses show that the nonlinear change in RT does not distort the partwise MVPA effects.

Moreover, in our analyses we did not exclude trials that followed timeout feedback (i.e. where the participant responded too slow in the previous trial). The timeout feedback might temporarily shift the focus to speed rather than accuracy so that subsequent errors might occur because the participant tries to meet the response time limit. However, we consider it as a benefit that our task captures a variety of error types, e.g. errors due to speed pressure (which can occur with or without previous timeout feedback), errors due to a failed response selection process, or premature responses. This relates more closely to errors committed in everyday life. In the same note, we also did not exclude trials where multiple responses were executed (which were most prevalent in non-signalled errors). The execution of multiple responses might influence error processing mechanisms because it impedes a clear identification of the response that was performed first, thus hindering the assessment of whether or not this response deviates from the correct response. However, on a conceptual level it is difficult to separate the execution of multiple responses from non-signalled errors, as a great portion of errors might only be non-signalled because of the multiple responses. Thus, excluding trials with multiple responses would limit the representativeness of our results, as we would ignore an important cause for non-signalled errors.

Finally, we opted for a constant order where the no-self-evaluation condition was always presented first. This results in a possible confound between the time on task and an effect of self-evaluation. We deliberately restrained from counterbalancing the two self-evaluation conditions, because we suspected an interaction between time on task and self-evaluation, in which case counterbalancing is not helpful to disentangle time on task from other effects (Hutchison et al., 1999; Keppel, 1982; Sayette et al., 2010). A carryover effect from the self-evaluation condition to the no-self-evaluation condition appeared highly plausible. If explicitly instructing participants to evaluate their response accuracy enhances the attentional resources allocated to this process (Grützmann et al., 2014), and if we had instructed participants to do so in the first part of the experiment, it is likely that this enhancement would have transferred to the second part of the experiment, even if the rating was not present. This is problematic, because we do not expect such a carryover effect when the no-self-evaluation condition precedes. Therefore, counterbalancing would interact with the order of conditions. Many researchers have precluded or at least warned about the implementation of counterbalancing whenever differential carryover effects are presumed, because they impede the interpretability of results (Keppel, 1982; Sayette et al., 2010; Winer, 1971). Previous studies that implemented counterbalancing or trial-by-trial variations risked capturing these differential carryover effects which stayed unanalysed in most of the studies. Our design precludes differential carryover effects, albeit at costs of possible time on task effects. By implementing several MVPA, we were still able to uncover an effect of self-evaluation that exceeds a time on task effect on brain activity patterns. We discuss the effects of time on task and the issue of counterbalancing more deeply in the Supplement.

Conclusion

In this study, we have implemented the recently developed 8ART. This has enabled us to demonstrate that in a more complex task with higher cognitive load, neural mechanisms of error processing vary with error detection. This is similar to the findings in the literature that used binary response tasks. We successfully replicated Stahl et al.'s (2020) findings regarding behavioural parameters and the later stages of error processing. We also observed the common $N_{e/c}$ effect (in contrast to Stahl et al., 2020), presumably due to the extended learning phase in our study. These results encourage further studies to implement more complex tasks that more accurately reflect decision-making processes in everyday life. We did not observe that self-evaluation enhanced early error processing mechanisms by increasing error significance. Instead, the absorption of attentional resources by the rating procedure might be responsible for the decrease in early error processing mechanisms we observed. Self-evaluation enhanced later error processing mechanisms, which presumably reflect error evidence accumulation. While early error processing might have been decreased due to the withdrawal of attentional resources, we assume that error evidence accumulation was exactly the process that absorbed the attentional resources, because it is crucial to successfully detect response accuracy. Beyond the influence of self-evaluation, we uncovered changes in earlier and broader processing mechanisms developing with time on task. These findings support the notion that a more holistic approach that exhausts information to a larger extent can complement traditional univariate approaches by further disentangling variations in brain activity patterns. Together, the different methodological approaches deepened our insight into the relation of self-evaluation and error processing in our complex task.

Study 2

Motor Inhibition versus Interference Suppression: Neural and Behavioural Features of Action Monitoring and Error Processing

Abstract

Motor inhibition and interference suppression are essential inhibitory mechanisms to master daily routines. When interference suppression fails, we become distracted by irrelevant stimuli. When motor inhibition fails, we cannot override a strong response impulse. In both cases, error processing is crucial to learn from our mistakes. We systematically investigated the antecedences and consequences of inhibition errors in the event-related potential over the entire trial course of a complex choice task. We assessed motor inhibition and interference suppression in separate conditions while using similar stimuli and identical instructions. For motor inhibition errors, we found evidence for an enhanced P2 (associated with deficient sensory gating) and an enhanced P3 (associated with resourceful conflict evaluation) compared to correct responses. For interference suppression errors, these neural correlates were more similar to their correct counterparts. Results from the lateralized readiness potential showed a prematurely initiated motor execution process for motor inhibition errors and a deficient response selection process for interference suppression errors. After the response, the error types displayed similar N_e components (associated with early error monitoring). The P_e (associated with error evidence accumulation) tended to be enhanced for motor inhibition errors, suggesting that an early inhibitory failure might be more evident than deficiencies during complex response selection. Together, our findings indicate that motor inhibition errors arise from a deficient early inhibitory process at the levels of perception and motor execution, whereas interference suppression errors arise during a complex response selection process where conflict processing is resourceful even for correct responses.

Introduction

Action inhibition and error commission are inherent parts of everyday life. Inhibition is an essential mechanism to master daily routines. When inhibition fails, error processing is crucial to learn from the mistake and to reach the goal in the next attempt. Investigating errors that result from inhibitory failures generated important insights into neurocognitive variations of error processing with the experimental context (e.g. task complexity, Maier et al., 2010; reward and punishment, Maruo et al., 2016, 2017; feedback, Olvet & Hajcak, 2009b), individual differences (e.g. narcissism, Mück et al., 2023; perfectionism, Stahl et al., 2015), and even clinical symptoms (e.g. OCD, Riesel et al., 2023; anxiety, Riesel et al., 2019). These studies used different tasks (e.g. Eriksen-Flanker task, Go/NoGo task) with high inhibitory demands to evoke errors. However, it is crucial to consider that inhibition is not a unitary construct, but comprises at least two facets: interference suppression and motor inhibition (Friedman & Miyake, 2004; Nigg, 2000). When *interference suppression* is deficient, we become distracted by irrelevant stimuli and lose focus of the target (e.g. when a television in the background distracts us from listening to a conversation), whereas in deficient *motor inhibition* a strong response impulse cannot be inhibited and one fails to override a prepotent response tendency (e.g. when we press the send button of an email although we just realised the attachment is missing). The aforementioned error-related findings are often based on the investigation of either motor inhibition errors or interference suppression errors. Thus, to some extent, these findings might be restricted to a specific error type. Importantly, although there is evidence for a correlation of the neural error processing markers of motor inhibition and interference suppression, there might be vital differences between the two errors types (Riesel et al., 2013), affecting error processing mechanisms differentially. Interestingly, despite their central roles, to our knowledge interference suppression errors and motor inhibition errors were never compared directly and systematically within a single error processing study. A systematic and detailed comparison of motor inhibition errors and interference suppression errors from stimulus onset to post-response adaptation will provide future studies a base to investigate their research questions (e.g. variations in error processing with the experimental context, individual differences and clinical disorders), which might be further affected by the task choice and the resulting error type. Thus, in a first step, the goal of the current study is to enable a direct comparison between motor inhibition errors and interference suppression errors by overcoming previous task-related restrictions (for details, see below). This allows us, in a next step, to systematically investigate variations in the processes before both types of inhibition errors arise (such as perception, attention, inhibition, response evaluation and preparation) and after their occurrences (such as detection processes and adaptive mechanisms).

Pre- and Post-Response Indicators Assessing the Time Course of Inhibition

Cognitive neuroscience methods such as the event-related potential (ERP) in combination with behavioural measures help to monitor the processing course from stimulus onset over error processing to adaptive mechanisms of the two inhibition errors. The ERP components that occur within the first 200 ms, such as the P1 (Luck et al., 1990; Mangun et al., 1993), the N1 (Haider et al., 1964; Luck et al., 1990) and the P2 (Bedoin et al., 2019; Benikos et al., 2013; Lijffijt et al., 2009), are usually used as markers of early stimulus processing and attention-related processes.

After the first 200 ms, the components N2 and P3 and the lateralised readiness potential (LRP, Coles, 1989; Gratton et al., 1988) are important markers of higher order decision processes and response preparation. The frontocentral N2 amplitude signals the need for motor inhibition, interference suppression, response conflict detection and conflict monitoring (Falkenstein, 2006; Gehring et al., 1992; Kok et al., 2004; Kopp et al., 1996), whereas the subsequent frontocentral P3 amplitude is associated with evaluation of the inhibition outcome (Kok et al., 2004; Ramautar et al., 2004) and response cancellation (Bedoin et al., 2019). The P3 latency is often used as a marker of stimulus evaluation time (Polich, 2007). The LRP is another helpful time-critical marker that splits the response time (RT) into the pre-decision time (i.e. from stimulus onset to onset of activity of the contralateral primary motor cortex) and motor processing (i.e. from activation of the primary motor cortex to onset of the overt motor response; Bryce et al., 2011; Nayak et al., 2020; Vuillier et al., 2016).

After response execution, action monitoring is reflected by the error negativity (N_e or ERN; Falkenstein et al., 1991; Gehring et al., 1993), which is associated with error processing and conflict monitoring (Yeung et al., 2004), and the error positivity (P_e; Falkenstein et al., 1991), which is discussed as an indicator of aware error processing and error evidence accumulation (Steinhauser & Yeung, 2010, 2012). Both components are typically more pronounced for errors than for correct responses (N_c or CRN, P_c; Falkenstein et al., 1991; Gehring et al., 1993).

On the behavioural level, adaptational mechanisms subsume post-error slowing (PES; Laming, 1968; Rabbitt & Rodgers, 1977) and post-error improvement of accuracy (Danielmeier & Ullsperger, 2011; Klein et al., 2007).

Inhibition Tasks: Characteristics and Challenges Investigating Inhibition Errors

To properly compare the course of the processing mechanisms of both error types using the above-mentioned markers, we have to address several task-specific characteristics. Interference suppression is usually induced by tasks such as the Eriksen-Flanker task (Eriksen & Eriksen, 1974) and the Stroop task (Stroop, 1935), whereas motor inhibition is usually evoked by tasks such as the Go/NoGo task (O’Connell et al., 2009; O’Keeffe et al., 2007; Shalgi et al., 2007; van Driel et al., 2012) and the Stop-Signal task

(Bender et al., 2016; Eagle et al., 2008; Logan & Cowan, 1984). These inhibition errors are highly reliant on the features of these task (see also the task-impurity problem; Burgess, 1997). In interference suppression tasks, a response is usually executed on every trial and errors are induced by stimulus incongruence. In contrast, motor inhibition tasks usually request withholding a response on certain trials without presenting (perceptually) interfering stimuli, thus provoking erroneous responses by lowering the motor threshold (i.e. increasing the tendency to respond). Trials where interference suppression is needed can provoke two types of overt responses – correct responses or errors. Trials where motor inhibition is needed provoke only one motor response, which is on error trials (commission errors), whereas correct trials require the suppression of an overt response (correct withhold). All these aspects particularly challenge contrasting the two error types, especially because motor inhibition errors lack their physically equivalent counterparts (i.e. overt correct responses on trials where inhibition is needed).

Previous Findings on the Comparison of Interference Suppression and Motor Inhibition

Some studies implemented so-called hybrid tasks, where both prepotent response tendencies and incongruence were induced to contrast motor inhibition and interference suppression (Bedoin et al., 2019; Bissett et al., 2009; Blasi et al., 2006; Brydges et al., 2012, 2013; Groom & Cragg, 2015; Vuillier et al., 2016). Most of these studies investigated successful inhibition processes and thus predominantly focus on contrasting motor inhibition and interference suppression in the context of correct responses. Using a hybrid task, Brydges et al. (2012) found that the N2 peaked later for interference suppression than for response inhibition. This indicated that motor inhibition and interference suppression can be dissociated on a neural level and that motor inhibition might occur before interference suppression. In contrast, Vuillier et al. (2016) reported no such differences in N2 latencies. However, both studies found a more negative N2 for trials where no interference suppression was needed (congruent trials) compared to trials where it was needed (incongruent trials). Vuillier et al. (2016) stated that a larger N2 is more reliably linked to motor inhibition than to interference suppression. Bedoin et al. (2019) reported increased N2 and P3 amplitudes on NoGo trials (motor inhibition) compared to Go trials (no motor inhibition) when interference suppression was effortless but only an increased P3 amplitude when interference suppression was effortful. The authors interpret the N2 and the P3 as response restraint and response cancellation, respectively, and assume that, on NoGo trials with effortful interference suppression, high-level forms of response restraint failed (reflected in the absence of an N2 increase) due to the resourcefulness of interference suppression. Studies using the LRP hypothesized that interference suppression is part of response selection, whereas motor inhibition is part of response execution (Bender et al., 2016; Bissett et al., 2009; Nee et al., 2007; Sebastian et al., 2012). Together, these studies form important steps towards a deeper understanding of the two inhibition facets and their relationship.

However, hybrid tasks cannot fully rule out the challenges of the traditional inhibition tasks when it comes to disentangling motor inhibition and interference suppression, especially for the different error trials. For example, in hybrid tasks, both conditions are often combined within trials (e.g. Bedoin et al., 2019; Vuillier et al., 2016) or the conditions are mixed between trials within blocks (e.g. Bissett et al., 2009; Blasi et al., 2006). While combining the conditions allows for an investigation of interaction effects between the two inhibition facets, mixing them implies that participants need additional attentional resources to switch between the different inhibitory mechanisms and that the inhibitory processes become less distinct. Our newly developed task aims at resolving some of these caveats. With this in mind and considering neural and behavioural measures, we systematically investigate how motor inhibition errors and interference suppression errors relate to each other in their antecedences and consequences.

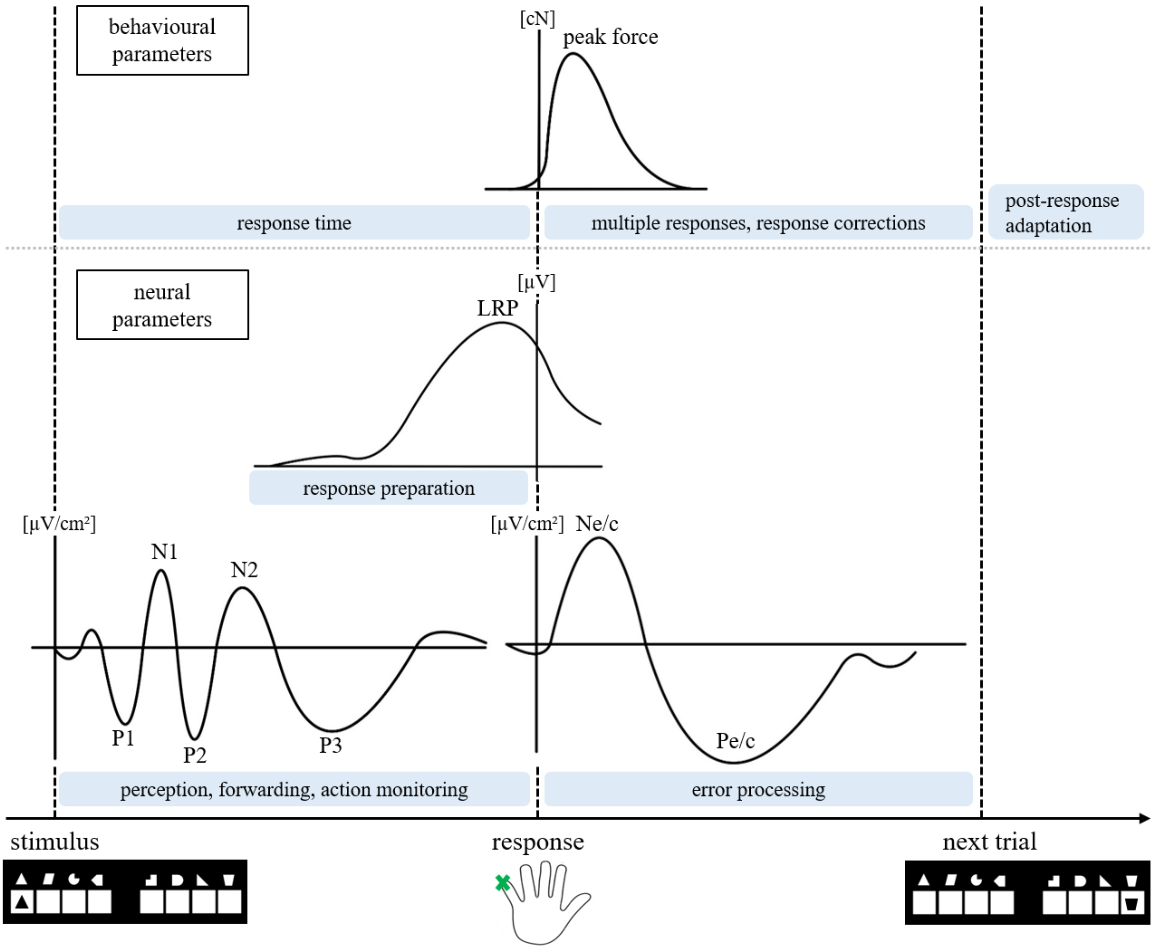
Study Objectives

Most of the above-mentioned studies used binary tasks, which restricts generalizability to more complex everyday actions. Thus, the foundation of our task is a modified version of the eight-alternative response task (8ART; Stahl et al., 2020; see also Porth et al., 2022), which is a speeded choice task with a complex response selection. In the modified task, we operationalized motor inhibition and interference suppression in two separate conditions (i.e. without combining them within trials or mixing them on a trial-by-trial basis) while using similar stimuli and the same task instruction (for details see Methods). In the motor inhibition condition, we manipulated the frequency of the congruency between the stimulus identity (response defining feature) and the stimulus location (irrelevant feature). The more frequent congruent trials had the purpose to induce a prepotent response tendency towards stimulus location, which needed to be overridden during the less frequent incongruent trials. Overt correct responses on the less frequent trials form the counterpart to motor inhibition errors, that is missing in typical Go/NoGo tasks. In the interference suppression condition, we used distractor stimuli to evoke interference on the level of stimulus processing and response selection.

To this end, we systematically investigated and contrasted the antecedences and consequences of the two inhibition errors, starting at stimulus processing and ending with post-response adaptation. Our goal was to elucidate the cascade of cognitive processes underlying the two error types. This delivers insights into the degree to which the two error types are dissociable, allowing future studies to better understand their findings by considering different error types. We assessed neural and behavioural parameters corresponding to different stages of the processing stream from stimulus presentation to post-response adaptation for a comprehensive overview on the cognitive processing underlying the two error types. The processing stream and the associated variables are depicted in Figure 2-1.

Figure 2-1

Processing stream from stimulus onset to post-response adaptation



Note. Overview on the assessed behavioural measures (upper panel) and neural parameters including an illustration of their typical event-related potential waveforms and associated cognitive processes (lower panel) from stimulus onset to the next trial.

On a neural level, correlates of early perceptual stimulus processing (reflected by the P1 and N1) form the beginning of the processing stream. These neural correlates could serve as first potential markers of the dissociations between the two error types. The perceptual information is forwarded (a process often associated with the P2) to higher processing mechanisms such as action/conflict monitoring and conflict evaluation (associated with the N2 and P3). Meanwhile, the timing and course of preparatory patterns can be observed in the LRP. After response execution, error processing mechanisms are reflected by the N_e and P_e. On the behavioural level, response times capture the process from stimulus onset to response execution. During the time window of the response, response force and the percentage of multiple responses are often interpreted as error-specific indicators for uncertainty and online inhibition (Porth et al., 2022; Stahl et al., 2020). As measures of post-response adaptation we assessed PES and the post-response accuracy which often display modulations that are specific to different error types (Damaso et al., 2020).

The behavioural and neural parameters we assess in the current study have been equally related to motor inhibition and interference suppression in the past. Studies that jointly investigated motor inhibition and interference suppression predominantly focused on correct responses and implemented hybrid tasks, which complicate a clear distinction between the two inhibition types. This and considering the novelty of our task prevented us from making predictions on potential differences between motor inhibition errors and interference suppression errors. Hence, our study was exploratory and served the purpose of generating first insights into these potential differences.

Methods

Participants

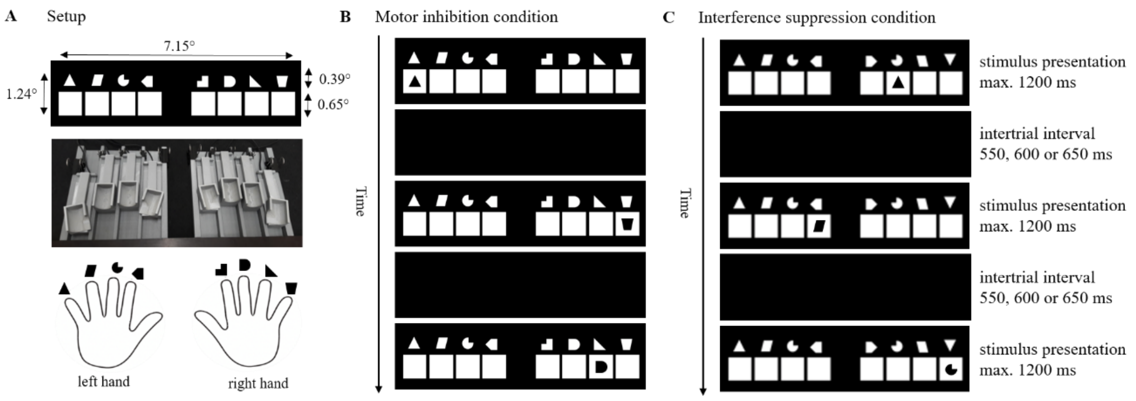
We collected data from 41 participants. We had to exclude datasets from several participants due to technical issues (defect light sensor: $n = 6$), insufficient data quality (not enough trials left after artefact rejection: $n = 3$) and insufficient number of trials of at least one response type (at least six trials necessary for error-related ERP components: $n = 2$; see Olvet & Hajcak, 2009a). The final sample comprised 30 undergraduate psychology students (9 males, 19 females, 2 diverse; age: $M = 23.9$ years, $SD = 6.0$ years). This sample size allowed uncovering effect sizes of $\eta_p^2 \geq .22$ including interactions, given a Type I error of .05 and a power of .80 (MorePower, Campbell & Thompson, 2012). We used an online system (Gotzhein & Elson, 2020) for participant recruitment. The participants were rewarded with course credits and reported normal or corrected-to-normal vision. The ethical board of the Faculty of Humanities at the University of Cologne approved the study and written consent was collected from each participant.

Procedure

We implemented a modified version of the 8ART (Porth et al., 2022; Stahl et al., 2020), where participants responded to appearing stimuli according to an instructed stimulus–response assignment in a given response time window (Figure 2-2). The task consisted of several trials, in each of which eight white squares were presented in the centre of a black screen. A different shape was located above each square. The participants placed their fingers, excluding their thumbs, on eight force-sensitive response keys. The response keys were assigned to the shapes above the squares. During each trial, one of the shapes also appeared in one of the eight white squares. The participants were instructed to respond to the appearing shape with the finger it was assigned to as fast and as accurately as possible. For this, they were given a response time limit of 1000 ms. As soon as the participants responded with one of the eight response keys, the trial was finished, and the next trial started. The inter-trial interval randomly varied between 550, 600 and 650 ms. If a response exceeded the response time limit, the feedback ‘even faster’ (German: *noch schneller*) appeared on the screen. No other feedback was given to the participants.

Figure 2-2

Setup of experiment and task design



Note. (A) Setup of the experiment: stimulus presentation including visual angles, force-sensitive response keys and stimulus–response assignment; (B) exemplary trials from the motor inhibition condition: The first two trials induce a prepotent response tendency (location matches stimulus identity) and in the third trial the location indicates an incorrect response tendency that has to be inhibited; (C) exemplary trials from the interference suppression condition with a target stimulus and a mirrored distractor stimulus.

The experiment consisted of two within-subject conditions. The first condition aimed at evoking prepotent response tendencies (*motor inhibition condition*). In this condition, the shapes appeared more frequently in the white square beneath their matching shape (congruent trials) with the purpose of creating an association between response and stimulus location. Every three to five trials, the appearing shape was not presented beneath its matching shape (incongruent trials). A retrieval of previously encoded associations between response and stimulus location would lead to an erroneous response. Instead, participants had to ignore this association and respond to the stimulus identity instead. Thus, motor inhibition is required when the more frequent stimulus-response assignments create prepotent response tendencies that need to be overridden on the less frequent trials (exemplary trials are depicted in Figure 2-2 B). The condition consisted of 10 blocks with 80 trials each, leading to a total of 800 trials for this condition. Short breaks between blocks ensured time to rest for the participants.

The second condition aimed at evoking responses based on distractor stimuli (*interference suppression condition*). This condition included the four shapes of the left side from the motor inhibition condition which were mirrored regarding their spatial orientation, leading to a set of eight shapes that resembled each other, serving as the interference-inducing distractor stimuli. We always mirrored the same four symbols to avoid differential effects of shifting or re-learning between the conditions. The appearance of the shapes inside the white squares varied randomly in location, with all shape–location combinations having the same probability. However, the shapes never appeared beneath their matching shape (as this would decrease the need for interference suppression) or their individual counterpart (as this would permit differentiating responses guided by distractor influence from responses guided by stimulus location). Interference suppression is required when the distractors induce conflict (e.g. regarding stimulus processing or response selection), which needs to be resolved in order to respond correctly (exemplary trials are depicted in Figure 2-2 C). This condition consisted of six blocks with 80 trials each, leading to a total of 480 trials for this condition.

In total, the motor inhibition condition contained more trials than the interference suppression condition because the congruent trials were used to generate prepotent response tendencies, but were not of interest for the analyses as no inhibition was required. Each inhibition condition was split into two parts and the order of conditions was balanced between participants (ABAB/BABA design) to avoid sequence effects. Each part started with two short practicing blocks consisting of 20 trials where participants were given the opportunity to get used to the task set and the response time limit. The total duration of the experiment was approximately 45 minutes, including breaks.

Apparatus

Figure 2A shows the set of stimuli and their visual angles. The physical centre of the screen is marked by the black square in the middle between the two rows of four white squares. A chin rest was used to constrain excessive head movements and to keep the screen distance (88 cm) constant. A luminance-sensitive photosensor was mounted on the bottom corner of the screen to assess real-time stimulus onsets. The eight response keys measured response force via embedded force sensors (FCC221-0010-L, DigiKey MSP6948-ND). A VarioLab AD converter digitized the response signal at a sampling rate of 1024 Hz. When the sensors registered the exceedance of 40 cN on one of the response keys, the keypress was registered as a response. The response keys were calibrated to the weight of the participant’s fingers and adjusted to their length.

Electrophysiological Data

The electroencephalogram (EEG) signal was recorded with 63 active Ag/AgCl electrodes (Acticap, Brain Products) that were placed onto the scalp according to the standard international 10–20 system (Jasper, 1958): FP1, FP2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C3', C1, Cz, C2, C4, C4', C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, PO9, O1, Oz, O2, PO10. An electrode placed on the left mastoid served as an online reference. The averaged signal from the left and right mastoids was used for offline re-referencing. The electrooculogram (EOG) was recorded with four passive bipolar Ag/AgCl electrodes (ExG-Amplifier, Brain Products) positioned horizontally next to the eyes and vertically above and below the left eye. A BrainAmp DC amplifier (Brain Products) recorded the EEG and EOG signals continuously at a sampling rate of 500 Hz and a filter from DC to 70 Hz.

We applied a 50 Hz notch filter to clean the data from line noise. For the stimulus-locked ERP components, we locked the trial intervals onto the stimulus onset and segmented time windows from –100 ms to 700 ms. For the response-locked components, we locked the trial intervals onto the response onset and segmented time windows from –100 to 500 ms. We performed baseline corrections based on the 100 ms before stimulus onset or response onset, respectively. We corrected for eye blinks with an ocular correction algorithm (Gratton et al., 1983) and rejected artefacts where ERP waves exceeded ± 100 µV. To clean each electrode's activity from the activity of nearby electrodes and to detach the signal from the reference electrode we carried out a current source density (CSD) analysis (Perrin et al., 1989).

The assessment specifics of the investigated ERP components are listed in Table 2-1. The time windows and electrode sites for the ERP components were determined based on literature (P1 and N1: Luck et al., 1990; Boksem et al., 2005; P2: Lijffijt et al., 2009; N2 and P3: Bokura et al., 2001; $N_{e/c}$ and $P_{e/c}$: Gehring et al., 2012). Analogously to Boksem et al. (2005), the P1, N1 and P2 were measured at the electrode sites where the grand average waveforms showed the local maxima. For the N2 and P3, after visual inspection of the individual and grand average waveforms the onsets of the time windows were preponed by 50 and 100 ms, respectively. The majority of studies we used to determine electrode sites and time windows for the ERP components did not implement CSD-transformations. However, studies using CSD-transformation report similar time windows and electrode sites (e.g. Riesel et al., 2013; Stahl et al., 2020). We quantified the mean amplitudes as ± 2 data points around the peak on the participant level separately for each condition. The averaged activity from five data points decreases the influence of noise opposed to a simple peak measure (Clayson et al., 2013) and the narrow time window for the sharp components with distinct peaks avoided covering activity from preceding and subsequent components. For the slower components (P3, $P_{e/c}$) we additionally averaged activity from 40 data points (see Supplement).

The LRP is defined as the average of the two differences [C4´(t) – C3´(t)] for the left hand and [C3´(t) – C4´(t)] for the right hand (Coles, 1989), using the jack-knifing method to estimate LRP onsets (Miller et al., 1998; Stahl & Gibbons, 2004; Ulrich & Miller, 2001) and transforming the jack-knifed onset values according to Smulders (2010). While the duration of stimulus processing and response selection can be assessed by examining the time from stimulus onset to LRP onset, the duration of response activation and motor processes can be assessed by examining the time from LRP onset to response onset (Smulders & Miller, 2011). We used a 50% criterion to determine the LRP onset for both the stimulus-locked and the response-locked LRP (i.e. the onset was determined at the first time point where the LRP exceeded 50% of the peak amplitude for at least five consecutive time points). To enable comparisons between correct responses and errors, we based LRP computations for errors on the hands with which the responses were executed instead of the hands with which the correct response would have been executed. For the LRP data we applied a 30 Hz low-pass filter.

Table 2-1

Information on the assessed event-related potential components: time windows, electrodes, quantifications, the events they refer to and current source density analyses

Component	Time window	Electrode	Quantification	Event	CSD
P1	90 to 160 ms	PO7, PO8 [a]	Mean amplitude [b]	Stimulus-locked	Yes
N1	150 to 210 ms	P1, P2	Mean amplitude	Stimulus-locked	Yes
P2	150 to 250 ms	PO9, PO10	Mean amplitude	Stimulus-locked	Yes
N2	150 to 300 ms	FCz	Mean amplitude	Stimulus-locked	Yes
P3	200 to 500 ms	FCz	Mean amplitude	Stimulus-locked	No
$N_{e/c}$	0 to 150 ms	FCz	Mean amplitude	Response-locked	Yes
$P_{e/c}$	100 to 300 ms	Cz	Mean amplitude	Response-locked	Yes
LRP [c]	–100 to 800 ms	C3´, C4´	LRP onset	Stimulus-locked	No
LRP	–500 to 200 ms	C3´, C4´	LRP onset	Response-locked	No

[a] When the signal was not distributed centrally, we averaged activity from both electrode sites where the signal was largest.
[b] Mean amplitude as ± 2 data points around the peak.
[c] LRP = lateralised readiness potential

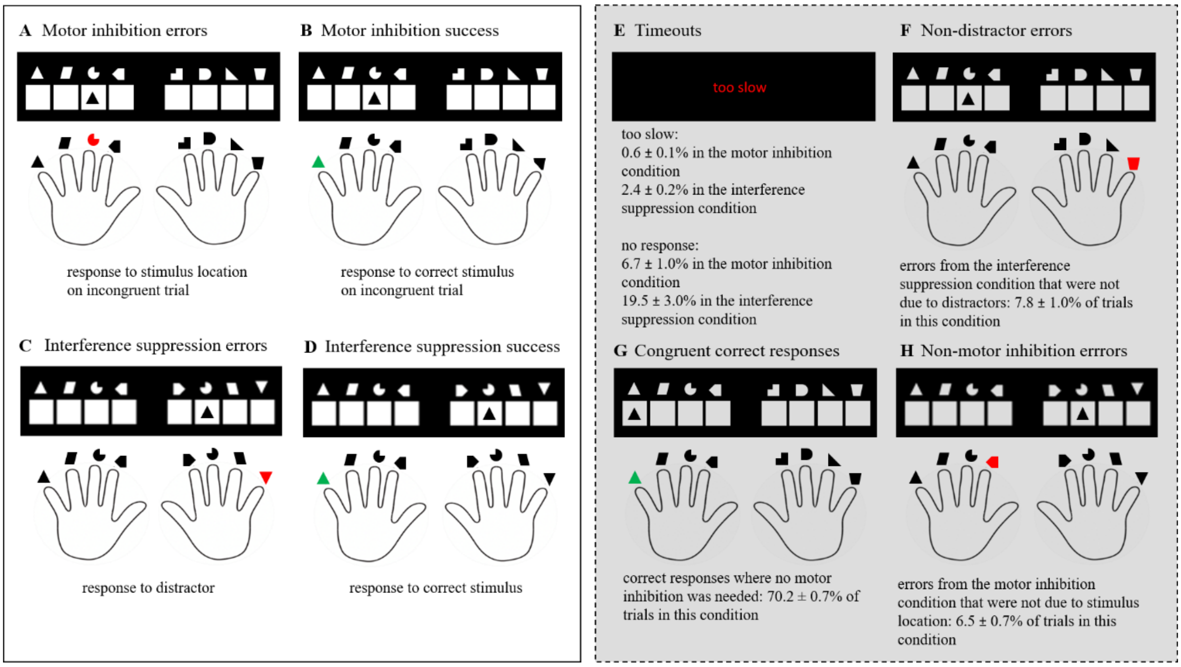
We report the results of CSD-transformed data for all ERP components except for the P3. We conducted the CSD analyses to obtain a signal that is cleared from activities of neighbouring electrodes (Luck, 2014). For the P3, however, visual inspection revealed that this component was quite distorted by the CSD analyses. This can be explained by the fact that the P3 emerges from activity across the entire midline of the scalp, where it is interpreted in different ways depending on where it is maximal. The topographies (see Results) confirmed that an increasing activity across the midline seems to be an inherent feature of the component in our task. Therefore, for the P3 we report results of the untransformed data (see Discussion for more information on the P3 in our study).

Statistical Analyses

Response Types. We included four response types in our analyses: motor inhibition errors, motor inhibition success, interference suppression errors, and interference suppression success. We defined motor inhibition errors as incongruent trials in the motor inhibition condition where participants erroneously responded to the location at which the shape appeared (i.e. according to the prepotent response tendency). Motor inhibition success refers to incongruent trials from the motor inhibition condition where the correct response was selected despite the mismatch between stimulus location and response. Hence, motor inhibition errors and motor inhibition success are both based on incongruent trials where response tendencies elicited by stimulus location potentially need to be ignored. In other words, congruent trials on which the stimulus location matched the response location were not considered in the analyses, regardless of their outcome (error, correct), as there was no incongruent information that required inhibition. We defined *interference suppression errors* as erroneous trials from the interference suppression condition where participants responded to the counterpart of the appearing shape (i.e. the distractor). *Interference suppression success* subsumed all correct responses from this condition as all trials implied a potential need for interference suppression. Figure 2-3 gives an overview over the four response types and the excluded trials.

Figure 2-3

Response types and excluded trials



Note. The left panel shows examples for (A) a motor inhibition error, (B) motor inhibition success, (C) an interference suppression error, and (D) interference suppression success. The right panel shows examples for excluded trials that are (E) too slow responses and trials where no response was given, (F) errors from the interference suppression condition that were not responses to the counterpart of the appearing shape (i.e. the distractor), (G) correct responses from the motor inhibition condition on congruent trials, and (H) errors from the motor inhibition condition that where not responses to stimulus location.

Behavioural Analyses. We analysed median RT, the percentage of multiple responses and the peak force with a series of 2×2 repeated-measures ANOVAs with the factors *inhibition type* (motor inhibition vs. interference suppression) and *accuracy* (success vs. error). Post hoc tests (Tukey’s HSD for within-comparisons) were implemented for significant ANOVA interactions. RT was defined as the time interval from stimulus onset to the exceedance of a force of 40 cN on one of the response keys. Multiple responses were defined as trials where the response force exceeded 40 cN on more than one key. We contrasted the percentage of response corrections of motor inhibition errors and interference suppression errors with a two-tailed *t*-test.

We computed PES by subtracting the averaged RT of trials preceded by motor inhibition/interference suppression success from the averaged RT of trials preceded by motor inhibition/interference suppression errors. As the visual inspection of our data did not indicate a confound of error frequencies and response time level, there was no need to quantify PES as a difference measure of pre-error trials and post-error trials (Dutilh et al., 2012; for details on the appropriateness of PES quantifications see Pfister & Foerster, 2022). Additionally, based on the recommendations of Pfister and Foerster (2022) we computed pre-error speeding with the same method to assess whether motor inhibition errors are preceded by more speeding (than interference suppression errors) as the preceding series of congruent trials should lower the motor threshold in this condition. We analysed whether motor inhibition errors and interference suppression errors showed PES and pre-error speeding that differed significantly from zero (one-sided *t*-tests) and whether the two error types differed from of each other regarding PES and pre-error speeding (two-sided *t*-tests).

We contrasted the post-response improvement of accuracy between motor inhibition errors and motor inhibition success and between interference suppression errors and interference suppression success (two-tailed *t*-tests). We did not compare the post-response improvement of accuracy between inhibition types because in the motor inhibition condition the probability of a subsequent correct response was inherently higher.

Electrophysiological Analyses. We analysed all ERP parameters with a series of 2x2 repeated-measures ANOVAs with the factors *inhibition type* (motor inhibition vs. interference suppression) and *accuracy* (success vs. error). Post hoc tests (Tukey’s HSD for within-comparisons) were implemented for significant ANOVA interactions.

Depicting cognitive processes from stimulus presentation to post-response adaptation implies conducting a series of inference statistical tests. One possibility to deal with this multitude of tests is to adjust the alpha level, ensuring a global Type-I error rate of 5% for the entire study. While the issue of multiple comparison conduction is certainly of importance (for ERP research see Luck & Gaspelin, 2017), such a strict way of testing conflicts with the exploratory purpose of this study. As the main goal was to generate first insights into the differences between motor inhibition errors and interference suppression errors, we report the uncorrected statistics in the following, but naturally replication studies are needed.

Results

Behavioural Results

The descriptive statistics for the behavioural parameters are presented in Table 2-2 separately for each response type.

Table 2-2

Means and standard error of means for the behavioural parameters for each inhibition type (motor inhibition, interference suppression) and accuracy (success, errors)

	Motor inhibition				**Interference suppression**			
	Success		**Error**		**Success**		**Error**	
	M	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
Response rates[a] [%]	73.3	3.1	26.7	3.1	81.9	1.7	18.1	1.6
Response time [ms]	754.6	11.2	538.7	9.5	765.8	8.7	788.3	9.3
Multiple responses [%]	2.8	0.9	3.2	0.9	3.9	1.2	5.0	0.8
Response force [cN]	164.9	11.8	111.9	6.3	189.8	15.6	159.2	12.2
PES [ms]	-	-	11.5	6.3	-	-	29.4	5.4
Pre-error speeding [ms]	-	-	–18.0	5.0	-	-	–7.3	4.4
Post-response acc. [%]	93.9	0.8	92.6	1.5	59.3	2.8	57.8	3.2

[a] Computation is based solely on the trials included in the analyses. For response rates of the excluded trials see Figure 2-3.

The ANOVA for RT revealed a significant main effect of accuracy, $F(1, 29) = 260.02$, $p < .001$, $\eta_p^2 = .90$, and of inhibition type, $F(1, 29) = 572.56$, $p < .001$,

$\eta_p{}^2 = .95$. The interaction of accuracy and inhibition type was also significant, $F(1, 29) = 432.23$, $p < .001$, $\eta_p^2 = .94$. Post hoc tests indicated that motor inhibition errors had the shortest RTs, all p values < .001, followed by the motor inhibition success and interference suppression success, which did not differ significantly from each other, $p = .529$. The RTs for interference suppression errors were longer than for motor inhibition success and interference suppression success, $p < .001$ and $p = .006$, respectively.

For the percentage of multiple responses there was no significant main effect of accuracy, $F(1, 29) = 0.73$, $p = .401$, $\eta_p^2 = .02$, but there was a significant main effect of inhibition type, $F(1, 29) = 24.15$, $p < .001$, $\eta_p^2 = .45$, indicating more multiple responses in the interference suppression condition (8.9 ± 1.7%) compared to the motor inhibition condition (6.0 ± 1.6%). The interaction of accuracy and inhibition type was not significant, $F(1, 29) = 0.42$, $p = .521$, $\eta_p^2 = .01$.

The analyses of response corrections showed that there were more response corrections for interference suppression errors (3.2 ± 0.6%) than for motor inhibition errors (1.3 ± 0.4%), $t(29) = –2.92$, $p = .007$, $d = –0.53$.

For response force, there was a significant main effect of accuracy, $F(1, 29) = 52.53$, $p < .001$, $\eta_p^2 = .64$, a significant main effect of inhibition type, $F(1, 29) = 20.82$, $p < .001$, $\eta_p^2 = .42$, and a significant interaction of accuracy and inhibition type, $F(1, 29) = 8.55$, $p = .007$, $\eta_p^2 = .23$. Response force was highest for interference suppression success, all p values < .022, and lowest for motor inhibition errors, all p values < .001. Motor inhibition success and interference suppression errors did not differ significantly, $p = .869$.

PES was shown for motor inhibition errors (11.5 ± 6.3 ms), $t(29) = 1.82$, $p = .039$, $d = 0.33$, and interference suppression errors (29.4 ± 5.4 ms), $t(29) = 5.49$, $p < .001$, $d = 1.00$, and both error types differed significantly from each other, $t(29) = 2.32$, $p = .027$, *d* = 0.42. Pre-error speeding was shown for motor inhibition errors (-18.0 ms ± 5.0), *t*(29) = -3.63, *p* < .001, *d* = -0.66., and as a (non-significant) tendency also for interference suppression errors (-7.3 ms ± 4.4), *t*(29) = -1.65, *p* = .055, *d* = -0.30, and both error types differed significantly from each other, *t*(29) = 2.53, *p* = .017, *d* = 0.46. A post-response improvement of accuracy was neither shown in the motor inhibition condition, *t*(29) = –0.91, *p* = .370, *d* = –0.17, nor in the interference suppression condition, *t*(29) = –1.08, *p* = .219, *d* = –0.20.

Electrophysiological Results

The descriptive statistics for the electrophysiological parameters are presented in Table 2-3. For the N2 and P3, a minimum number of 20 and 14 trials, respectively, is recommended to ensure internal consistency (Rietdijk et al., 2014). The analyses of the $N_{e/c}$ and $P_{e/c}$ should contain at least six error trials (Olvet & Hajcak, 2009a). On average, 30 motor inhibition errors (*SD* = 17, *min* = 9, *max* = 74) and 49 interference suppression errors (*SD* = 27, *min* = 12, *max* = 115) were included in the ERP analyses (here: response-locked) after artefact rejection, indicating sufficient stability of the ERP measures.

Table 2-3

Means and standard error of means for the electrophysiological parameters for each inhibition type (motor inhibition, interference suppression) and accuracy

	Motor inhibition				**Interference suppression**			
	Success		**Error**		**Success**		**Error**	
	M	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
P1 amplitude [µV/cm^2]	0.21	0.04	0.21	0.04	0.19	0.03	0.21	0.04
N1 amplitude [µV/cm^2]	–0.05	0.02	–0.03	0.02	–0.03	0.02	–0.04	0.02
P2 amplitude [µV/cm^2]	0.20	0.03	0.22	0.03	0.18	0.03	0.19	0.03
N2 amplitude [µV/cm^2]	–0.10	0.01	–0.10	0.01	–0.08	0.01	–0.10	0.02
P3 amplitude [µV]	2.3	0.7	3.7	0.7	3.1	0.6	3.5	0.7
P3 latency [ms]	306.3	12.0	303.5	11.9	331.5	13.9	315.3	9.2
$N_{e/c}$ amplitude [µV/cm^2]	–0.12	0.02	–0.14	0.02	–0.10	0.02	–0.14	0.02
$P_{e/c}$ amplitude [µV/cm^2]	0.07	0.03	0.19	0.03	0.04	0.02	0.11	0.02
S-LRP onset[a] [ms]	502.8	18.3	672.1	5.9[b]	582.0	16.6	394.7	11.0
R-LRP onset[c] [ms]	–140.2	9.6	–223.8	3.8	–114.3	7.1	–139.3	25.0

[a] S-LRP = stimulus-locked lateralized readiness potential

[b] The S-LRP onset for motor inhibition errors does not lie within the average RT (for details, see discussion).

[c] R-LRP = response-locked LRP

Early Stimulus Encoding: P1, N1 and P2

The ANOVA for the P1 amplitude showed no significant main effect of accuracy, $F(1, 29) = 3.35$, $p = .077$, $\eta_p^2 = .10$, no significant main effect of inhibition type, $F(1, 29) = 1.21$, $p = .280$, $\eta_p^2 = .04$, and no significant interaction of accuracy and inhibition type, $F(1, 29) = 0.43$, $p = .518$, $\eta_p^2 = .01$. For the N1 amplitude, there was no significant main effect of accuracy, $F(1, 29) = 0.55$, $p = .464$, $\eta_p^2 = .02$, or inhibition type, $F(1, 29) = 0.27$, $p = .604$, $\eta_p^2 < .01$, but a significant interaction of accuracy and inhibition type, $F(1, 29) = 6.35$, $p = .018$, $\eta_p^2 = .18$. However, post hoc tests did not reveal any significant differences between the response types, all *p* values > .234.

For the P2 amplitude, there was a significant main effect of accuracy, *F*(1, 29) = 6.40, *p* = .017, η_p^2 = .18, and of inhibition type, *F*(1, 29) = 5.53, *p* = .026, η_p^2 = .16, and a non-significant tendency for an accuracy-by-inhibition type interaction, *F*(1, 29) = 3.85, *p* = .059, $\eta_p{}^2$ = .12. Although this effect did not reach a 5% significance level, we cannot rule out that it was driven by a higher P2 amplitude for motor inhibition errors than for the other three response types (see descriptive statistics in Table 2-3).

Inhibition and Conflict Monitoring: N2 and P3

For the N2 amplitude, there was no significant main effect of accuracy, *F*(1, 29) = 3.03, *p* = .092, η_p^2 = .09, or of inhibition type, *F*(1, 29) = 1.68, *p* = .205, η_p^2 = .05, and no significant interaction of accuracy and inhibition type, *F*(1, 29) = 0.78, *p* = .384, η_p^2 = .03.

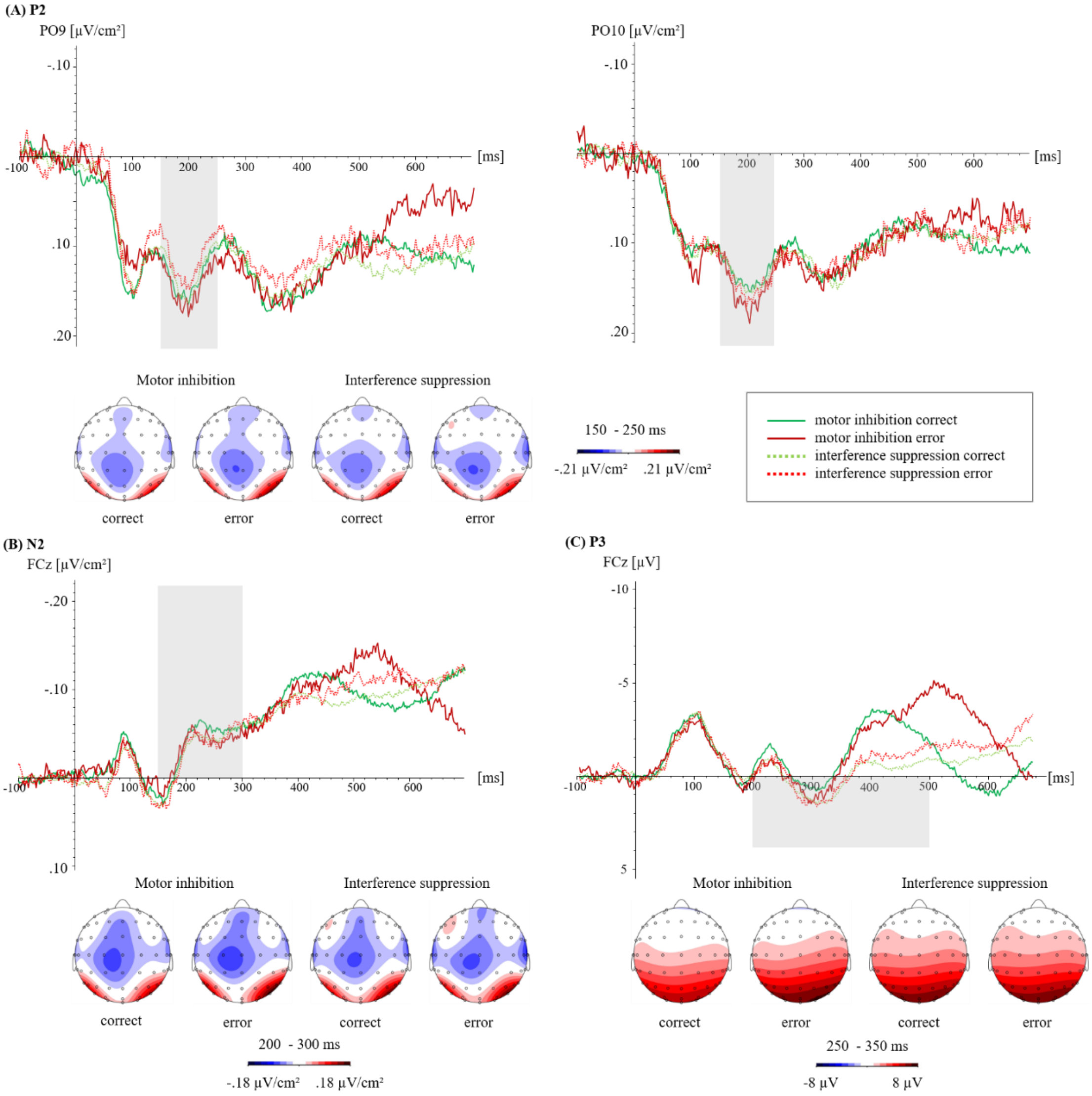
For the P3 amplitude, there was a non-significant tendency for an effect of accuracy, *F*(1, 29) = 3.78, *p* = .062, η_p^2 = .12. There was no significant main effect of inhibition type, *F*(1, 29) = 0.53, *p* = .471, η_p^2 = .02, but there was a non-significant tendency for an accuracy-by-inhibition type interaction, *F*(1, 29) = 3.73, *p* = .063, η_p^2 = .11. Although this effect did not reach a 5% significance level, we cannot rule out that it was driven by the lower P3 amplitude for motor inhibition success (see descriptive statistics in Table 2-3).

The P3 latencies did not vary significantly with accuracy, *F*(1, 29) = 1.12, *p* = .299, η_p^2 = .04, but there was a significant main effect of inhibition type, *F*(1, 29) = 4.21, *p* = .049, η_p^2 = .13, indicating longer latencies in the interference suppression condition. There was no significant interaction of accuracy and inhibition type, *F*(1, 29) = 0.40, *p* = .532, $\eta_p{}^2$ = .01.

The averaged ERP waveforms and the respective topographical maps of the P2, N2 and P3 are depicted in Figure 2-4 (Figures for paired differences of the waveforms for the P2, N2, and P3 can be found in the Supplement).

Figure 2-4

Averaged stimulus-locked ERP waveforms and topographical maps



Note. Averaged current source density transformed waveforms separately for each response type (A) of the P2, measured at PO9 and PO10 electrode sites as the mean amplitude ± 2 data points around the positive peak in the time window of 150–250 ms after stimulus onset; (B) of the N2, measured at FCz electrode as the mean amplitude ± 2 data points around the negative peak in the time window of 150–300 ms after stimulus onset. (C) Averaged untransformed waveforms of the P3, measured at FCz electrode as the mean amplitude ± 2 data points around the positive peak in the time window of 200–500 ms after stimulus onset separately for each response type.

From Stimulus Encoding to Motor Execution: LRP

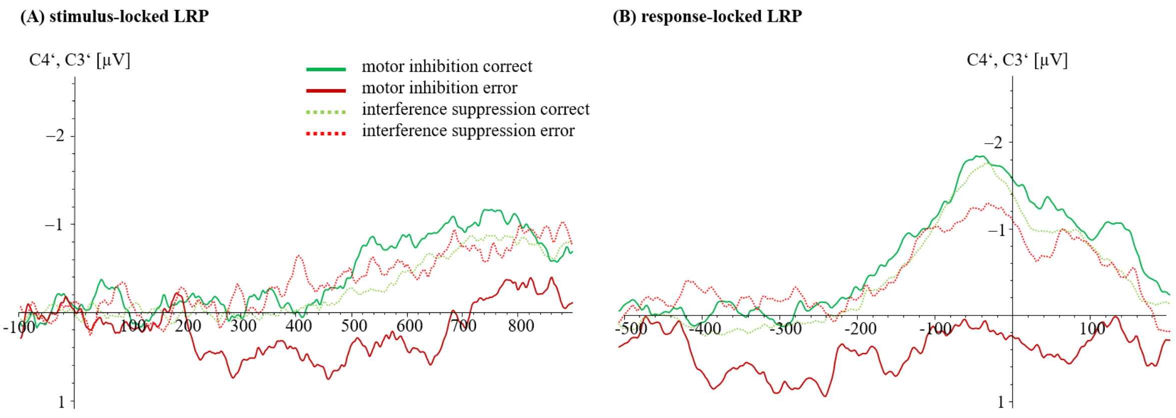
For the LRP analyses, we included datasets from 24 of 30 participants because the analyses required a minimum number of six trials *per hand*, which led to the exclusion of several participants due to an insufficient number of error trials. The ANOVA for the stimulus-locked LRP onset showed no significant main effect of accuracy, $F(1, 23) = 0.99$, $p = .330$, $\eta_p^2 = .04$, but a significant main effect of inhibition type, $F(1, 23) = 94.15$, $p < .001$, $\eta_p^2 = .80$, and a significant interaction of accuracy and inhibition type, $F(1, 23) = 330.47$, $p < .001$, $\eta_p^2 = .93$. Post hoc tests showed that the stimulus-locked LRP onset differed significantly between all response types, all p values < .001, with interference suppression errors having the earliest onset, followed by motor inhibition success, interference suppression success and then motor inhibition errors.

The ANOVA for the response-locked LRP onset showed a significant main effect of accuracy, $F(1, 23) = 14.58$, $p < .001$, $\eta_p^2 = .39$, and inhibition type, $F(1, 23) = 12.34$, $p = .002$, $\eta_p^2 = .35$, and a significant interaction of accuracy and inhibition type, $F(1, 23) = 5.78$, $p = .025$, $\eta_p^2 = .20$. Post hoc tests showed that the response-locked LRP onset was significantly earlier for motor inhibition errors than for the other three response types, all p values < .019, which did not differ from each other, all p values > .071.

The stimulus-locked and response-locked averaged LRP waveforms are depicted in Figure 2-5 (paired differences of the LRP onsets are depicted in the Supplement).

Figure 2-5

Averaged LRP waveforms



Note. Averaged lateralized readiness potential (LRP) waveforms derived from C3' and C4' electrode sites for each response type (A) stimulus-locked in the time window of –100 to 900 ms; (B) response-locked in the time window of –500 to 200 ms. To enable comparisons between successful responses and errors, we based LRP computations for errors on the hands with which the responses were executed.

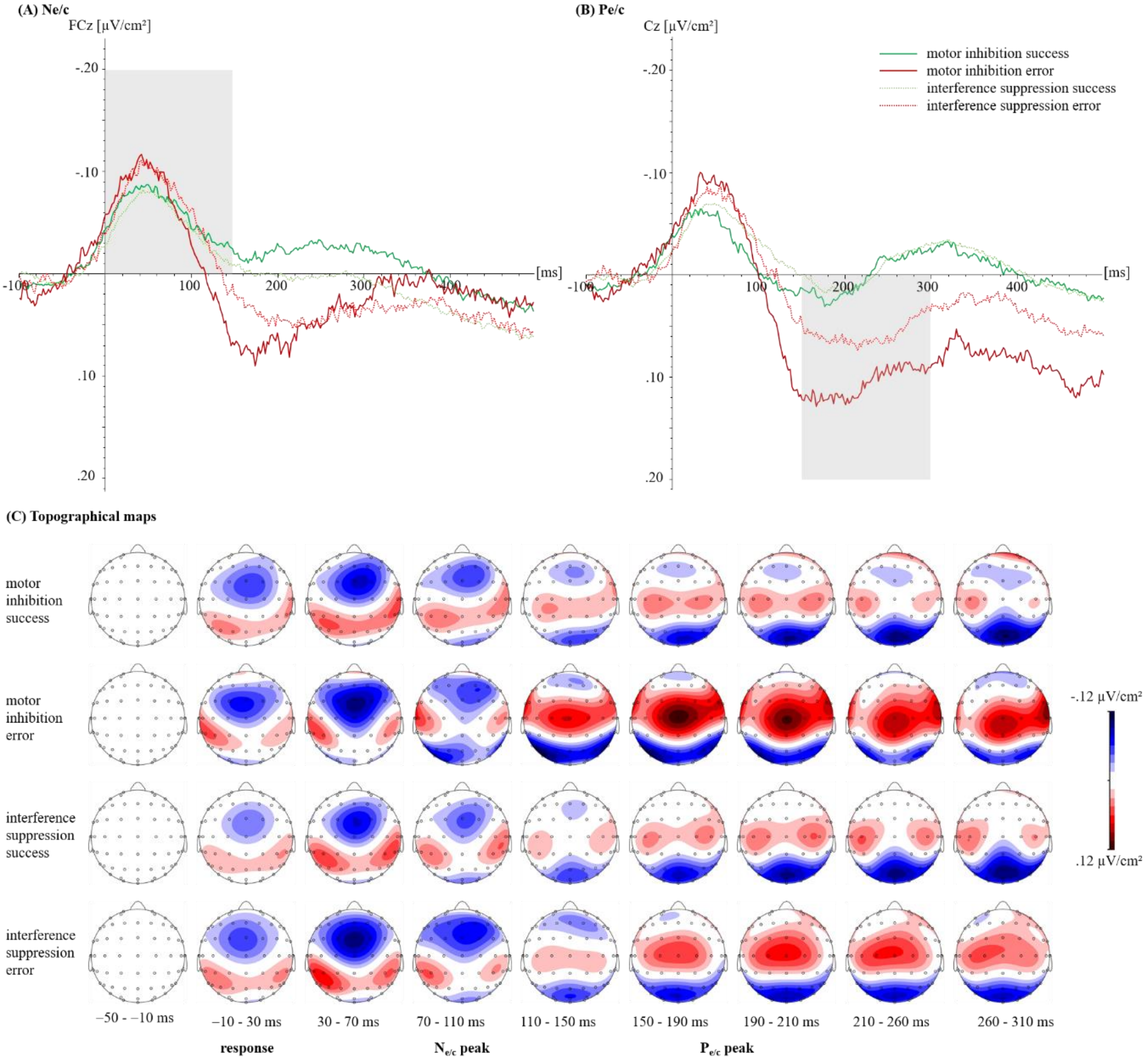
Error Processing: $N_{e/c}$ and $P_{e/c}$

For the $N_{e/c}$ amplitude, there was a significant main effect of accuracy, $F(1, 29) = 7.63$, $p = .010$, $\eta_p^2 = .21$, indicating higher amplitudes for errors than for correct responses. There was no significant main effect of inhibition type, $F(1, 29) = 1.33$, $p = .259$, $\eta_p^2 = .04$, and no significant interaction of accuracy and inhibition type, $F(1, 29) = 0.52$, $p = .478$, $\eta_p^2 = .02$. For the $P_{e/c}$ amplitude there was a significant main effect of accuracy, $F(1, 29) = 16.19$, $p < .001$, $\eta_p^2 = .36$, indicating higher amplitudes for errors than for successful responses, and a significant main effect of inhibition type, $F(1, 29) = 10.47$, $p = .003$, $\eta_p^2 = .27$, indicating higher amplitudes in the motor inhibition condition than in the interference suppression condition. There was a non-significant tendency for an accuracy-by-inhibition type interaction, $F(1, 29) = 3.91$, $p = .057$, $\eta_p^2 = .12$. Although this effect did not reach a 5% significance level, we cannot rule out that it was driven by the lower P_e amplitude for motor inhibition errors (see descriptive statistics in Table 2-3).

The CSD-transformed averaged ERP waveforms for the $N_{e/c}$ and $P_{e/c}$, together with the topographical maps, are depicted in Figure 2-6 (Figures for paired differences of the waveforms for the $N_{e/c}$ and $P_{e/c}$ can be found in the Supplement).

Figure 2-6

Averaged response-locked ERP waveforms and topographical maps



Note. Averaged current source density transformed waveforms separately for each response type (A) for the error negativity, measured at the FCz electrode as the mean amplitude ± 2 data points around the negative peak in the time window of 0–150 ms after response onset; (B) for the error positivity, measured at the Cz electrode as the mean amplitude ± 2 data points around the positive peak in the time window of 100–300 ms after response onset. (C) Topographical maps in the time windows of the $N_{e/c}$ and $P_{e/c}$.

Discussion

By using a modified version of the 8ART (Porth et al., 2022; Stahl et al., 2020), we created two conditions that represent two facets of inhibition: motor inhibition (by implementing rare stimulus combinations) and interference suppression (by implementing conflicting stimuli). Importantly, we aimed at keeping the perceptual differences between the two inhibition conditions at a minimum by using similar stimuli and instructions and reduced differences in motor execution by requiring a motor response on every trial. This was necessary to disentangle the response types psychologically rather than physically for the analyses following the Hillyard principle (Hillyard et al., 1973). In a first step, we systematically investigated how interference suppression and motor inhibition errors arise and how they differ in their antecedences. In summary, our results indicate that motor inhibition errors are based on the failure of an early inhibitory process that needs to be implemented *before* the initiation of a (complex) response selection, and that this type of inhibition does not proceed just on the level of motor execution. In contrast, interference suppression errors seem to arise *during* the (complex) response selection and seem relatively restricted to this process.

Before starting the discussion of the differing cognitive processes of the two error types, we emphasize that we do not postulate a one-to-one mapping of electrophysiological correlates and cognitive processes. In the literature, several different cognitive processes were discussed to underlie each ERP component. The following interpretations take place in the context of the literature closely related to our task types, which may simplify the interpretation at some levels, but avoid a complication on other levels.

Early Stimulus Encoding

Starting our comparison at stimulus onset, the two inhibition errors showed no significant differences in early attention-related perceptual encoding mechanisms reflected by the N1 and the P1. Interestingly, the two error types displayed variations already at the bridge between stimulus encoding and later processes. The enhanced P2 amplitude in the motor inhibition condition compared to the interference suppression condition is a first indicator of differences in the neural processing chain. The P2 is assumed to reflect an early inhibitory mechanism that prevents irrelevant stimulus features from being processed further in order to protect the system from initiating erroneous actions (i.e. sensory gating; see Bedoin et al., 2019; Benikos et al., 2013; Lijffijt et al., 2009), thus forming a bridge between early, low-level stimulus encoding and higher-order, more complex stimulus evaluation processes. Usually, a decreased P2 amplitude is associated with the success of this protective mechanism (Bedoin et al., 2019; Benikos et al., 2013; Lijffijt et al., 2009). For motor inhibition errors, there was a (non-significant) tendency for an increased P2 amplitude compared to successful responses, indicating that this protective mechanism might have failed. Thus, the irrelevant stimulus information (here, stimulus location) might have been processed further and was, in the end, responded to instead of ignored. For interference suppression, this difference between successful responses and errors appears smaller, suggesting that the P2-related early protective mechanism might be equally intact, which underlines that an early inhibitory mechanism might not be essential for this inhibition type.

Inhibition and Conflict Monitoring

The response types did not differ significantly regarding the N2 amplitude, thus response conflict (Nieuwenhuis et al., 2001; Yeung et al., 2004) and inhibition monitoring (Kopp et al., 1996) seem to proceed similarly in all four response types. Interestingly, in the motor inhibition condition, there was a (non-significant) tendency for a lower frontocentral P3 amplitude for successful responses than for errors, which indicates that the processes reflected by the P3 might be relevant for this error type. In the past, the P3 has been associated with various processes such as context updating (Donchin, 1981), attentional resource allocation (Kok, 2001), and response inhibition (Falkenstein et al., 1991). The scalp distribution of the P3 is often described as the amplitude change over the midline electrodes that increases in magnitude from frontal to parietal electrode sites (Johnson, 1993). Processes of context updating and resource allocation are associated with a more parietal maximum of the component, while inhibitory processes are accompanied by a more frontocentral P3 (Eimer, 1993; Polich, 2007).

In our study, the topographies show positive activity that increases along the midline electrodes from frontocentral to parietal sites. Additionally, there is a strong positive activity at occipital sites. The topographies from stimulus presentation to response onset (Figure S2-1 the Supplement) show that this positive occipital activity persists over the entire trial course. This suggests that in the time window of the P3 there might be overlapping activities from occipital sites and the P3, which increases from frontocentral to parietal sites. As the P3 seems to be parietally driven, we analysed differences between response types at Pz electrode site (for statistics, see Supplement) and found a lower P3 amplitude for motor inhibition success compared to the other three response types. At parietal sites, the P3 is often associated with attentional resource allocation (Polich, 2007). While a lowered frontocentral P3 amplitude for successful motor inhibition is somewhat unexpected as previous literature often associated inhibition success with increased P3 amplitudes (Bekker et al., 2005; de Jong et al., 1990; Dimoska et al., 2003; Falkenstein et al., 2000; Kok, 2001), a lowered parietal P3 amplitude related to attentional resource allocation is more in line with expectations. When a prepotent response tendency is overcome successfully, evaluation of the inhibition outcome is straightforward and needs less attentional resources, whereas the outcome evaluation is more resourceful when the initial inhibition fails. This again underlines the central role of an early inhibitory mechanism in the motor inhibition condition because, when it fails, it impedes the evaluation of the inhibition outcome during later processing.

In contrast, an early inhibitory mechanism is not useful in the interference suppression condition, because here the stimuli need to be processed thoroughly to resolve conflict on the stimulus level and during the complex response selection process. The high resemblance between targets and distractors presumably makes conflict monitoring and evaluation of the inhibition outcome equally resourceful for both errors and correct responses. This is also supported by the longer P3 latencies for responses from the interference suppression condition, which might indicate a longer stimulus evaluation process (see Polich, 2007), as well as the higher number of multiple responses and response corrections compared to the motor inhibition condition.

Motor Execution versus Response Selection

Even for eight alternative responses, the stimulus- and response-locked LRP patterns show distinct curves with clearly identifiable onsets for interference suppression errors and successful responses from both inhibition types. In contrast, for motor inhibition errors there is no distinct curve and the estimated onset of the stimulus-locked LRP lies after the mean response onset. The method was successful in the other three conditions, which indicates that motor inhibition errors are special and that the reasons for the lack of a clear single LRP onset are conceptual rather than methodological. Indeed, the LRP of motor inhibition errors is marked by multiple small LRP-like curves in the averaged signal, where a clear single onset is difficult to define even from visual inspection of the grand average. The positive deflection of the curve indicates preparation of response of the opposite response side. The small LRP-shaped curves might be an indicator of weakened response impulses, suggesting that motor inhibition did take place but was not sufficient to prevent the execution of a response. In line with this, response force was lowest for motor inhibition errors, underlining that inhibition on a motor level was strongly at play in this condition. Interference suppression errors, on the contrary, seem to arise from an insufficient response selection process and from the premature response to a distractor stimulus, reflected by the earlier stimulus-locked LRP onset for interference suppression errors compared to successful interference suppression.

Underpinning the interpretation of the P2 results, the LRP results for motor inhibition errors suggest an *early* failure of inhibition and a prematurely initiated motor execution process. For interference suppression errors, the LRP results point to a deficient response selection process and an unimpeded motor execution. This is in line with previous literature that associated interference suppression with response selection and motor inhibition with response execution (Bender et al., 2016; Bissett et al., 2009; Nee et al., 2007; Sebastian et al., 2012).

Error Monitoring and Error Evidence Accumulation

To elucidate similarities and differences in error processing mechanisms, we also investigated the consequences of executing motor inhibition and interference suppression errors. After response execution, both early and later error processing (N_e/P_e) resembled common literature findings and showed the expected accuracy effects (Falkenstein et al., 1991; Gehring et al., 1993). The two types of inhibition errors did not differ significantly regarding the N_e amplitude, suggesting that early conflict-related error monitoring processes (Yeung et al., 2004) might proceed similarly for both error types. This is in line with results from a previous study by Riesel et al. (2013), who found that the convergent validity of the N_e across the Flanker task and the Go/NoGo task was reasonably high (r = .43), and even higher for the $N_{e/c}$ difference (ΔNe, r = .65). However, they also report a marginally larger ΔN_e for the Flanker task compared to the Go/NoGo task. The authors suggest that this might be an effect of task difficulty because lower error rates in the

Go/NoGo task indicate that the Flanker task was more difficult. In our study, we cannot observe a clear difference in task difficulty and the two inhibition conditions are more similar regarding psychometric measures, which might explain why we do not observe a significant N_e difference.

In a series of studies, Maier and colleagues contrasted interference suppression errors (there: flanker errors) to errors that have other causes such as speed pressure or response confusion (there: non-flanker errors) (Maier et al., 2011, 2012; Maier et al., 2019; Maier & Steinhauser, 2016). Their studies consistently showed that the N_e is more pronounced for interference suppression errors than for other errors which the authors interpret as an indicator of higher error significance (Maier et al., 2012; Maier & Steinhauser, 2016) and stronger post-error adjustments in selective attention (Maier et al., 2011). They argue that interference suppression errors have a higher error significance because they violate two task goals – responding accurately and suppressing the influence of distractors – whereas other errors only violate the first goal (Maier et al., 2012). They also showed that interference suppression errors are followed by stronger selective attention adjustments and discuss the N_e as a correlate of these adjustments (Maier et al., 2011). Analogously, motor inhibition errors might also violate two task goals – responding accurately and suppressing a prepotent response tendency. Hence, they might imply a similar error significance, explaining the similar N_e magnitude between the two error types. They might also imply a comparable need for adaptive adjustments because a previously correct response pattern (e.g. the association between response and stimulus location) led to an error and thus needs to be reviewed (Wiswede et al., 2013).

For later error processing, we found a (non-significant) tendency for a higher P_e amplitude for motor inhibition errors compared to interference suppression errors. This might suggest that a failure in motor inhibition is more evident than for a failure in interference suppression errors, where deficiencies occur during the more complex response selection. The higher number of multiple responses and response corrections for interference suppression errors might additionally hinder error processing for this error type. In this sense, in future studies it would be interesting to replicate this finding and to assess error detection via a trial-wise self-evaluation of response accuracy after each response to investigate whether error detection is indeed more accurate for motor inhibition errors. Interestingly, Riesel et al. (2013) did not find significant differences in P_e amplitude between a two-choice Go/NoGo task and a two-choice Flanker task. This again suggests that the smaller P_e for interference suppression errors in our study was due to the more complex response selection, an effect that does not seem to unfold in tasks with only two response alternatives.

Post-response Adaptation

Last, on the behavioural level, we examined the adaptational mechanisms that the two error types launched. PES can be interpreted as an indicator of an adaptive mechanism to avoid subsequent errors due to premature responding as well as an indicator of an orienting response towards an expectancy violating event (Wessel, 2018). Our results show less slowing after motor inhibition errors compared to interference suppression errors. Thus, for motor inhibition errors slowing might be less useful as an adaptive mechanism: As a motor inhibition error is always followed by a congruent trial where a response has to be made solely guided by response location, slowing to avoid another error might not be necessary. Furthermore, motor inhibition errors might elicit a weaker orienting response, as the source of the error might be easier to identify than for interference suppression errors, which is also reflected by the tendency for a higher P_e amplitude for motor inhibition errors. The higher pre-error speeding for motor inhibition errors suggests that the sequence of congruent trials preceding motor inhibition errors lowered the motor threshold and induced a tendency to respond according to stimulus location.

Types of Conflict in the 8ART

The 8ART involves at least three types of conflict: (1) the conflict between stimulus location and response location (analogously to a Simon task), (2) the conflict between target stimuli and distractor stimuli (analogously to a Flanker task), and (3) the motor conflict between responses, because responses are mapped onto targets and distractors (analogously to a Flanker task) or because a response that is triggered by previous stimulus-response association competes with other responses (analogously to a Go/NoGo task; for a detailed description of different types of conflict see Hommel, 2011).

A mismatch between stimulus location and response location (1) is present in both the motor inhibition condition and the interference suppression condition. In the motor inhibition condition this mismatch is less frequent than a match. This might increase the weight to which stimulus location is considered during stimulus processing and response selection (Hommel, 2011). Therefore, the incongruence on less frequent trials might conflict with the response tendency previously encountered during congruent trials. On the contrary, in the interference suppression condition, the stimulus location never matches the response location. Thus, stimulus location never predicts the response location which might decrease the weight of stimulus location for stimulus processing and response selection and the corresponding level of conflict. A conflict between target stimuli and distractor stimuli (2) is only present in the interference suppression condition where the stimuli resemble each other. In both conditions, different responses compete and elicit motor conflict (3): In the interference suppression condition distractors and targets are mapped onto different response keys. In the motor inhibition condition the response triggered by stimulus location competes with the other responses.

Although these different types of conflict play important roles in the two conditions, they might not always be present. In the motor inhibition condition, over the course of the task a deviation between stimulus location and response location every third to fifth trial might be anticipated after some practise. This might weaken the impact of stimulus location and lead to an adaptation of respective stimulus-response associations, decreasing the level of conflict and rendering motor inhibition less needed. In the interference suppression condition, the conflict between target and distractor might be reduced or eliminated on trials where the shape appears closer to the target than to the distractor. While for motor inhibition errors and interference errors it seems likely that at least one form of conflict was at play, not every successful response necessarily implies the presence of conflict and the need for inhibition in the same way or at all. Hence, it is important to note that successful motor inhibition and successful interference suppression refer to correct responses where the two types of inhibition might be causal for the successful response, but one cannot rule out that in some cases inhibition was not involved.

Limitations

With the modified 8ART, we aimed at addressing the task impurity problem of the two investigated inhibition errors to enable a structured, direct comparison between the two error types. Despite the many similarities, the two inhibition conditions did differ in some respects. First, the probability of the need for inhibition was different between conditions. While inhibition was potentially needed on every trial of the interference suppression condition because distractors were always present, inhibition was only needed on every third to fifth trial in the motor inhibition condition. The resulting differences in cognitive control strategies cause differences in the reliability of the measurement of N_e (Meyer et al., 2013). More precisely, Meyer et al. (2013) found the N_e amplitude to be more affected by the number of error trials in a Go/NoGo task (12 error trials required for a correlation of r = .80 of the N_e based on a subset of error trials with the grand average N_e) than in a Flanker task (8 trials required). This becomes problematic when the number of error trials included in the ERP analyses falls below the number that is required for sufficient reliability (Meyer et al., 2013). In our study, after the artefact rejection, all 30 datasets contained more than 8 interference suppression errors and 28 datasets contained more than 12 motor inhibition errors (2 datasets contained 11 and 9). Thus, we conclude that the different probabilities of the need for inhibition in the two inhibition conditions do not diminish the reliability of our ERP measures and we consider the differences in inhibition probability essential to evoke the mechanisms of interest.

Second, in the motor inhibition condition participants were confronted with congruent trials (match between stimulus location and response location) and incongruent trials (mismatch between stimulus location and response location), while there were only incongruent trials in the interference suppression condition. We did not include congruent trials in this condition as they imply a different level of interference suppression and might be thus less likely to generate interference suppression errors.

Third, the stimuli in the two conditions are similar but not identical, leading to perceptual differences between conditions. To evoke interference suppression errors, the shapes had to be more similar to each other than in the motor inhibition condition. This was crucial to provoke interference suppression errors, which is why it poses a perceptual difference that could not be resolved.

To systematically investigate variations in processing from perception to post-response adaptation of two inhibition errors, we had to compute multiple comparisons. Although we assume that each performed test targets a different research question, we cannot ensure that they are entirely independent, which would impact the false discovery rate. Consequently, while the alpha level for each individual test amounts to five percent, the global Type-I error rate, i.e. the probability that one of the effects is a false positive, would be considerably larger. It is important to underline that our investigations were exploratory: we implemented a newly developed complex response task, created two settings tailored to investigate inhibition errors and did not derive a priori assumptions regarding differences between the two error types. Therefore, while our study delivered important findings and serves as a solid basis for further investigations, replicational studies are needed.

Conclusion

In our study, we elucidated the neural and behavioural antecedences and consequences of motor inhibition errors and interference suppression errors in a complex choice task. Our findings suggest that motor inhibition errors arise from a deficient early inhibitory process at the levels of perception and motor execution that precedes the start of a more complex response selection process. Early attention-related processing mechanisms seem intact for motor inhibition errors. On the bridge to later, more elaborate processing, sensory gating might be the first process that is deficient, thus information that cues the prepotent response tendency is not suppressed but processed further, and conflict evaluation becomes more resourceful. In the end, this early failure of perceptual and motor inhibition might lead to a more pronounced error evidence accumulation process. In contrast, our findings indicate that interference suppression errors arise during the complex response selection process. Along the entire processing stream, they appear more similar to their correct counterparts. Early attention-related processes and sensory gating seem equally intact for errors and correct responses, and conflict evaluation seems equally resourceful. Interference suppression errors are accompanied by a high number of multiple responses and response corrections, which together with the more complex response selection process presumably might lead to less error evidence accumulation.

All in all, the results from our systematic analyses show that our complex choice task is suitable to investigate and contrast interference suppression errors and motor inhibition errors. We explored the processing mechanisms of both error types from stimulus processing to post-response adaptation. Naturally, replicational studies are needed and the relationship between the two error types can be investigated further, for example, by addressing error detection and decision confidence. As our approach allows a direct comparison of the two types of inhibition errors, future studies will be able to explore variations in the processing mechanisms with traits related to inhibition such as impulsivity and selective attention capacity.

General Discussion

Decades of research underline that error commission and error processing are multifarious constructs (Stahl et al., 2020). Investigating them with binary response tasks has generated refined insights into underlying basic cognitive processes and their contextual variations (e.g. Bedoin et al., 2019; Grützmann et al., 2014; Nieuwenhuis et al., 2001). Now, to understand the constructs more profoundly, error commission and error processing need to be studied in more complex choices. A complex task, however, implies that various processes coincide and are even more susceptible to contextual variations. A proper way to disentangle these different processes is a multimethod approach, where cognitive dynamics are investigated systematically from various angles using different measures and analysing techniques. Therefore, in our two studies, we implemented the 8ART as a complex choice task, and systematically investigated variations in neural and behavioural correlates of error commission and error processing using different parameters and analyses.

Advantages of the Eight-Alternative Response Task

The complex stimulus-response mapping of the 8ART allows us to capture conflict on the stimulus level, the response level, and the level of stimulus-response assignments within a single task. This renders the 8ART highly adaptive: By varying the frequency of stimulus-response combinations, we were able to manipulate the need for motor inhibition, and by implementing distractors, we were able to induce the need for interference suppression. By adding a detection rating with similar configurations as the main task we were able to assess additional processes without the need to reduce task complexity. Our studies demonstrate that based on these modifications different paradigms can easily be incorporated in the 8ART. The task is built on the principles of a Simon Task (i.e. conflict elicited by an incongruence between stimulus and response location), and our second study shows that it can also comprise the principles of a Go/NoGo task (i.e. conflict elicited by a prepotent but inadequate response tendency) and a Flanker task (i.e. conflict elicited by distractors). Analogously, in future work, other principles can be integrated through slight modifications. For example, the principles of a Stroop task (i.e. conflict elicited by a prominent but irrelevant perceptual information, Stroop, 1935) can be implemented by presenting the eight stimuli in different colours and by manipulating the congruency between the colour of the appearing shape and the colour of the correct target. The benefit of combining high levels of conflict with challenging stimulus-response assignments is underlined in our second study, where the additional challenges of motor inhibition and interference suppression led to a high number of errors, which is always of interest in EPR research (Olvet & Hajcak, 2009a).

All in all, our two studies exemplify that the 8ART serves as a useful tool to extend and exceed binary response tasks. Implementing the 8ART allowed us to replicate well-established effects related to error processing and to uncover error-related variations in neural and behavioural features that go beyond findings from binary choices. This underlines that the 8ART is well-suited to investigate error commission and error processing in complex choices. Of course, despite the increase of task complexity in the 8ART, to this end this only forms a first step towards depicting everyday life decisions. For a closer approximation, even more complex tasks are necessary. For example, simulating a game of chess or playing a piece on the piano (Proverbio et al., 2017; Ruiz et al., 2009) might be further alternatives to approximate error commission in everyday life, while still assuring high-quality EEG recordings in the laboratory.

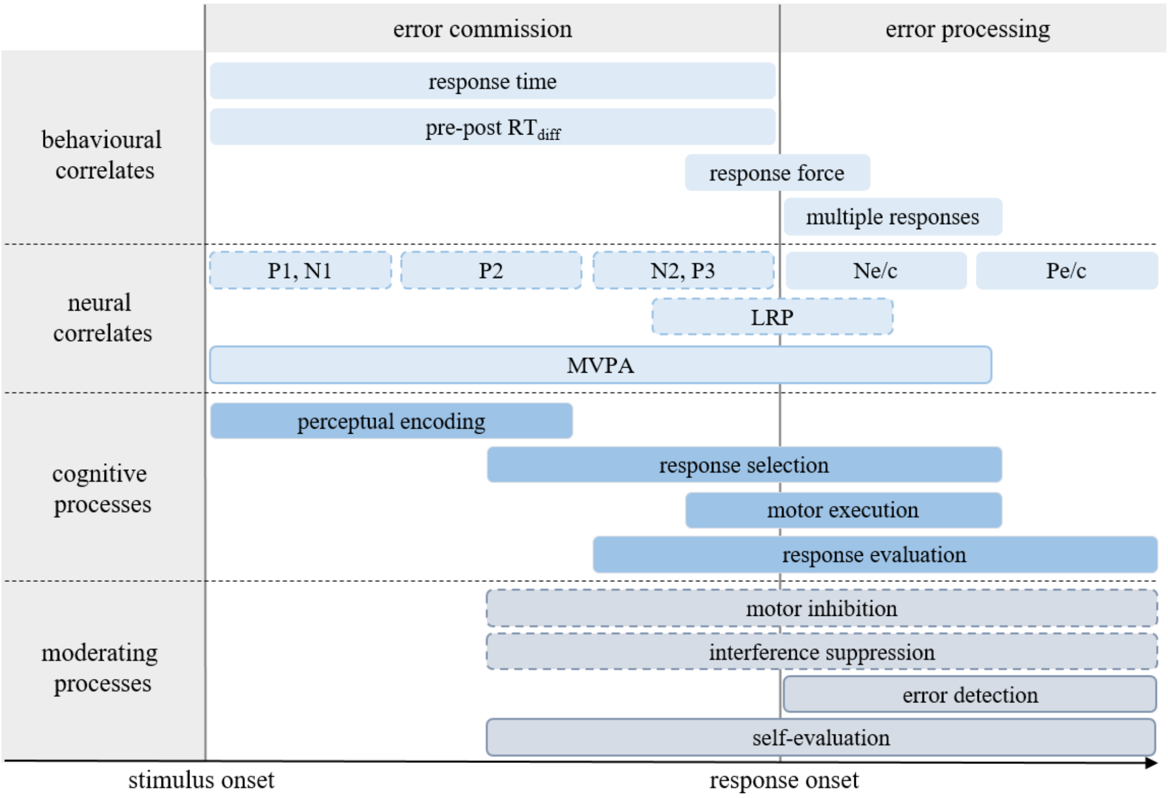
Despite the advantages of the 8ART, the implementation of the task is accompanied by certain challenges. The complexity of the task arises from the complex stimulus-response mapping and with that, different levels of conflict (e.g. between stimulus identity and response location, between motor responses, and between stimuli). Consequently, different processes intermix and can be difficult to disentangle. Thus, to distinguish these coinciding processes, the neural and behavioural correlates of error processing need to be elucidated systematically and with different methods. Therefore, in our two studies, we implemented traditional ERP analyses, machine learning techniques and a systematic analysing approach. The following abstracts encapsulate our core findings as well as further research questions that can be derived from our observations. In this context, additional methodological approaches that will help to answer these questions in future research are introduced.

A Multimethod Approach to Error Commission and Error Processing

In our first study, we combined traditional ERP analyses with MVPA to assess variations in neural and behavioural correlates of error processing with error detection and self-evaluation. This delivered insights into how these mechanisms, that were already linked to error processing in binary choices (e.g. Grützmann et al., 2014; Nieuwenhuis et al., 2001), are associated with error processing in a more complex task. In our second study, we implemented a systematic analysing approach to investigate variations in error-related processes with motor inhibition and interference suppression from stimulus onset to post-response adaptation. Together, both of our studies pursued the endeavour of mental chronometry research, i.e. to establish the dynamics of cognitive processes systematically, in the field of error processing by joining behavioural parameters and neurocognitive measures. As an overview, the neural and behavioural parameters that we assessed in our two studies as well as the associated cognitive processes and moderating processes are illustrated in Figure 1.

Figure 1

Overview on our multimethod approach of study 1 and study 2



Note. Overview on the behavioural and neural correlates of error commission (left panel) and error processing (right panel) assessed in both studies (solid frame: only study 1, dashed frame: only study 2) as well as associated cognitive processes and moderating processes. This figure is used for illustration purposes and does not include specific assumptions about the timing and degree of parallelism of the underlying cognitive processes.

Neural Correlates of Error Processing in Complex Choices

Traditional ERP Analyses. By implementing traditional ERP analyses in our first study, we replicated common patterns of the $N_{e/c}$ related to error detection and found variations in the $P_{e/c}$ with error awareness and self-evaluation that exceed findings from binary response tasks. In our second study, we directly compared motor inhibition errors and interference suppression errors by assessing them in two separate but similar conditions of the same task, which allowed us to uncover differences in the $P_{e/c}$ between the two error types. Together, our findings suggest that analogously to binary response tasks, error detection, self-evaluation and action inhibition are crucial moderating processes that affect the shape of neural error processing correlates in more complex choices. Yet, our results indicate that these variations partly deviate from the patterns reported for binary response tasks, probably because the increased task complexity affects attentional resource allocation and response selection differently.

Machine Learning Techniques. Next to traditional ERP analyses, in our first study we showcased that machine learning techniques deliver valuable insights into the temporal dynamics of cognitive processing in complex choices. We found substantial differences in whole-brain activity between self-evaluation conditions, which point to an influence of self-evaluation on early and broader processing mechanisms related to attention or perceptual encoding. Our findings underline that a machine learning approach enables a more holistic investigation of cognitive dynamics (Haynes & Rees, 2006) that complements traditional ERP analyses, which is particularly beneficial in complex choices where a variety of cognitive processes intermix.

The implementation of MVPA in our first study presages the variety of opportunities that machine learning techniques provide for error processing research in future work. For example, investigating the temporal dynamics of conscious error detection by contrasting signalled errors and non-signalled errors will elaborate existing accounts that refer to error awareness as an explanation for observed $P_{e/c}$ but not $N_{e/c}$ variations (e.g. mismatch theory, error awareness account, see Introduction). More specifically, while variations with error awareness in the ERP often show exclusively for the $P_{e/c}$, MVPA results might indicate that brain activity patterns can be dissociated between signalled and non-signalled errors already at earlier time windows, possibly even before the response onset. If this was the case, it would be worthwhile to investigate the emergence of conscious error detection more systematically beginning at stimulus onset, analogously to the approach in our second study.

Pupil Dilation. In future work, our multimethod approach can be extended by further methods to obtain more insights into the cognitive dynamics of error detection and error significance. A parameter that has been linked to both variables is pupil dilation (Maier et al., 2019; Wessel et al., 2011), which can be measured via eye-tracking (Blascheck et al., 2014). In general, pupils enlarge with sympathetic activation, and constringe with parasympathetic activation (Steinhauer et al., 2004). An extensive body of research supports a higher pupil dilation for errors than for correct responses (e.g. Maier et al., 2019; Murphy et al., 2016; Rondeel et al., 2015). This might reflect an increased level of stress and emotional occupation with the error (Hajcak et al., 2004; Spunt et al., 2012; Bradley et al., 2008; Partala & Surakka, 2003), and an increased allocation of attentional resources to compensate capacity reductions due to negative affect and worrying (Moser et al., 2013; Rondeel et al., 2015). Error-related pupil dilation can be explained by projections from the ACC, which is thought to generate the $N_{e/c}$ and $P_{e/c}$, to the locus coeruleus, which in turn modulates cognitive control by projections to prefrontal regions (Aston-Jones & Cohen, 2005; Jepma & Nieuwenhuis, 2011).

Because pupil dilation has been shown to vary with error awareness (Wessel et al., 2011), it will be interesting to investigate whether pupil dilation reflects the same gradual pattern for signalled and non-signalled errors as the P_e in our complex task. Additionally, comparing pupil dilation between the two self-evaluation conditions will be particularly informative. In the self-evaluation condition, we assumed that attentional resources shift away from fast automatic error detection towards later error evidence accumulation. This should correspond to a decreased pupil dilation for errors immediately after the response in the time window of the N_e for this condition, and an increase in pupil dilation in the time window of the P_e. Moreover, if pupil dilation corresponds to cognitive effort (van der Wel & van Steenbergen, 2018) and attentional resource allocation (Rondeel et al., 2015), we can also expect differences between the two types of inhibition errors that we investigated in our second study. Specifically, the difference in pupil dilation between correct responses and errors should be smaller in the interference suppression condition, where response selection appears resource-intense independent of response accuracy, compared to the motor inhibition condition.

Computational Modelling. Another technique that can expand our multimethod approach to investigate error detection in future work is computational modelling. In a recent study we introduced the neuro-cognitive diffusion model, where we combined computational modelling of behavioural and neural parameters on the single trial level (Mattes et al., 2022). In general, the diffusion model is a prominent approach to approximate cognitive processes underlying binary decisions with the help of response times (Ratcliff, 1978, 1988). The model assumes that, beginning from a starting point (z_0), evidence for or against one of two response alternatives is accumulated continuously (drift rate v) until one of the two decision thresholds (a and 0) is reached and motor preparation starts (non-decision time t_0) (Ratcliff & McKoon, 2008). The drift rate is associated with the attention directed to the task (i.e. a higher drift rate equals a stronger focus on the task), and the decision threshold serves as an indicator for response caution (i.e. a higher threshold equals increased response caution) (Dutilh et al., 2012).

Although the diffusion model is predominantly used to depict binary decisions, it can still be applied in the frame of complex choices, for example to model error detection. Modelling error detection as a binary decision (signalled vs non-signalled error) generates information about the strength of evidence accumulation and the decision thresholds that are applied to the process of self-evaluation. This is particularly interesting when comparing different error types. For example, we could include a detection rating in both inhibition conditions of our task. If the higher P_e for motor inhibition errors compared to interference suppression errors indeed reflects facilitated error evidence accumulation, this should translate to the diffusion model parameters of the signalling response: the starting point (z_0) should shift towards the upper boundary (a, i.e. signalled error), and the decision thresholds (a and 0) should be closer together because less evidence needs to be accumulated until a decision is reached.

In sum, applying traditional ERP analyses, MVPA and a systematic analysing approach in our two studies generated valuable insights into variations of error processing mechanisms with error detection, self-evaluation and action inhibition. Employing further methods such as eye tracking and computational modelling will foster an even more profound understanding of error processing in complex choice tasks.

Neural Correlates of Error Commission in Complex Choices

Perceptual Encoding. In our second study, using traditional ERP analyses, we assessed early visual perception with the components P1 and N1, both occurring in the first 200 ms after stimulus onset (Haider et al., 1964; Luck et al., 1990; Mangun et al., 1993). There, we did not find significant error-specific variations related to motor inhibition and interference suppression. However, error-related variations in perceptual encoding might occur at later stages. To uncover these potential variations in future work, we need a different method. More specifically, assessing the visual fixation times and saccades via eye-tracking (Blascheck et al., 2014) would allow us to investigate whether visual processing in the 8ART continues throughout the entire trial course, whether it encompasses the entire stimulus set, and whether it is different for errors than for correct responses. For example, a vital source for interference suppression errors might be the increased visual attendance of distractor stimuli, while a feature of motor inhibition errors might be the lack of a targeted perceptual comparison between the appearing shape and the shape above. Conclusively, divergences in visual processing and perceptual encoding could pose vital sources for the two types of inhibition errors in our complex task that we can only uncover by adding a further method to our approach.

Response Selection. Next to perceptual encoding, response selection forms an important process in the cognitive cascade (Janczyk, 2017; Meyer et al., 1988). Findings from both of our studies accentuate that response selection and continuous response evaluation play a central role in our complex choice task. Response selection is especially complex in the 8ART as there are 64 possible combinations of stimulus identity and stimulus location, and the complexity of response selection is enhanced by the different levels of conflict (see above) that are present in this task. Traditional ERP analyses allowed us to investigate the P3 component as a marker for the allocation of attentional resources and the evaluation of conflict and inhibition (Polich, 2007) in our second study. We observed variations in the P3 with motor inhibition and interference suppression suggesting that the evaluation of conflict and inhibition is particularly resource-intense when response selection is challenging. The MVPA findings from our first study support the central roles of response selection and response evaluation in the 8ART. The self-evaluation condition was decodable from whole-brain activity patterns as early as 650 ms before the response onset, indicating that the two resource-intense evaluation ratings might have withdrawn attentional resources from response selection and evaluation during the main task. For future work it will be interesting to contrast motor inhibition errors and interference suppression errors with MVPA. When contrasting these two error types, the predominant role of response selection for interference suppression errors should result in a similarly early classification onset as for the two self-evaluation conditions.

Motor Execution. Besides response selection and evaluation, findings from both of our studies display variations in motor execution. In the first study, we investigated motor execution solely on the behavioural level. In the second study, traditional ERP analyses revealed variations in the LRP (as a marker for motor preparation, Coles, 1989; Gratton et al., 1988) related to motor inhibition errors. Adding to the multimethod approach, measuring the electromyogram (EMG) in future work will extend our insights into error-related variations of motor execution in complex choices. The EMG refers to the measurement of muscle activity as an indicator of peripheral motor processing (Stahl & Rammsayer, 2004). For EMG recordings, electrodes are placed on the ventral forearms (Miller & Hackley, 1992). Analogously to the EEG signal, continuous EMG recordings can be locked onto the response onset. The EMG peak amplitude marks the intensity of peripheral motor activity and the EMG latencies serve as an indicator for the speed of peripheral motor activation (Stahl & Rammsayer, 2004). Dissociating the temporal dynamics of motor activations is of particular interest for errors related to action inhibition. Our LRP findings suggest that in motor inhibition errors, early response inhibition is present but insufficient to prevent a response. These findings should transfer to the EMG, where we should find a lower intensity of peripheral motor activity for this error type. The speed of peripheral motor activation might either be accelerated in motor inhibition errors, reflecting the prematureness of a response, or elongated, reflecting a counteractive effect of inhibition.

Additionally, EMG recordings enable the investigation of partial errors. Partial errors refer to erroneous motor responses that have been initiated on the cortical level, but were inhibited just in time to prevent full motor execution (Burle et al., 2008; Resulaj et al., 2009). As partial errors imply the initiation of an incorrect response, they deliver insights into the amount of conflict elicited during response selection and into the strength of inhibitory counteraction. In the frame of our second study, it would be interesting to investigate whether smaller EMG onsets in favour of a distractor-based, erroneous response can be observed for successful interference suppression. These partial errors would suggest that the distractor stimuli interfered with the selection of the correct response even when interference suppression was successful in the end, supporting the notion that response selection in the interference suppression condition is particularly resource-intense independent of response accuracy. Unfortunately, measuring the EMG at the ventral forearms implies that the peripheral motor activation can only be dissociated between hands, not between single fingers. Thus, it is not possible to draw conclusions about the exact response that was initiated but not fully executed. However, in our task, the pairs of distractors are mapped onto different hands. Therefore, a partial error evoked with the hand opposite to the correct one suggests that the response initiation was influenced by a distractor, and an increased response force for the key corresponding to the distractor can deliver additional evidence.

Conclusively, by approaching error commission systematically and with different methods, we obtained information about visual processing, response selection, and motor execution in complex choice tasks. Future work can extend these findings by investigating perceptual encoding via eye tracking, response selection via machine learning techniques, and motor execution with the EMG.

Behavioural Correlates of Error Commission and Error Processing in Complex Choices

In both of our studies we assessed response times, response force, the percentage of multiple responses, and post-error slowing as behavioural correlates of error commission and error processing. Mental chronometry research demonstrates that behavioural correlates represent important indicators for variations in cognitive processes (Posner, 2005). Indeed, both of our studies displayed error-related variations in response force that deliver interesting insights into the cognitive processes associated with this parameter. The response force was lower for non-signalled errors compared to signalled errors in our first study, while it was higher for interference suppression errors compared to motor inhibition errors in our second study. The lowered response force for non-signalled errors can be explained by a higher uncertainty about the correctness of the response and the execution of multiple responses instead of a single distinct response. However, if uncertainty and multiple responses indeed caused this lower response force, we would also expect a lower response force for interference suppression errors, where the number of multiple responses was higher, and uncertainty about the correctness of the response should have been stronger than for motor inhibition errors. Yet, we found the opposite pattern. This suggests that potential force reductions due to uncertainty and multiple responses in interference suppression errors might have been outweighed by force reductions due to inhibition during response execution (Allain et al., 2004; Gehring & Knight, 2000) in motor inhibition errors.

As the mechanisms behind error-specific variations in response force are still subject for debate (Overhoff et al., 2022; Siswandari et al., 2019), it would be interesting to predict the response force from whole-brain activity patterns with MVPA (using support-vector regressions) in future work. More precisely, predicting the response force from brain activity patterns separately for motor inhibition errors and interference suppression errors might help to disentangle the differential influences of uncertainty and inhibition. For interference suppression errors, the prediction onset of response force might be earlier than for motor inhibition errors because uncertainty should arise already during response selection, while motor inhibition should influence response force closer to the response onset.

Another parameter that delivered interesting findings in both of our studies is post-error slowing. In our second study, we found more slowing after interference suppression errors compared to motor inhibition errors. However, we do not know whether the higher slowing for interference suppression errors reflects an orienting response (induced by difficulties in findings the source of an error), a more adaptive mechanism (i.e. a higher focus on the relevant features of the task), or a combination of both. To test these alternatives in future work, we can apply the diffusion model (Ratcliff, 1978, 1988). If the higher slowing reflects an orienting response, we should find a lower drift rate (less attention directed to the task) in the post-error trial. If it reflects an adaptive mechanism, the decision thresholds should be further apart (higher response caution) and we should observe a higher drift rate (more attention directed to the task). In a next step, future work can explore variations in post-error slowing more profoundly by relating it to the neural correlates of error processing. For example, it will be interesting to predict the diffusion model parameters of the post-error trial for both types of inhibition errors from the single trial $N_{e/c}$ and $P_{e/c}$ amplitudes of the error trial (analogously to Mattes et al., 2022). Of course, the diffusion model is more appropriate for modelling binary responses, but there are other models that can accommodate for the complexity of our task (e.g. the linear ballistic accumulation model; Brown & Heathcote, 2008). Similarly, it will be interesting to investigate the slowing we found after non-signalled errors in our first study, as we assumed that it reflects both an orienting response and adaptational mechanisms elicited by the previous error and the precedent erroneous signalling response. Thus, it will be particularly informative to predict the diffusion model parameters of the subsequent trial from the response times of the detection rating of the (non-signalled) error trial.

All in All, conflating the behavioural findings from both of our studies expands our knowledge about the cognitive processes associated with response force and post-response adaptation in the context of error commission and error processing. Extending our acquired knowledge by additional analysing techniques such as support vector regressions and the (neuro-cognitive) diffusion model in future research will allow us to further investigate these processes and to understand relations to other mechanism more deeply.

Conclusion

In the frame of two studies, we used a multimethod neurocognitive approach that allowed us to investigate the timing, dynamics and the architecture of cognitive processes related to error commission and error processing systematically and precisely. Our two studies promote the assessment of eclectic performance parameters (beyond response times) and the implementation of neurocognitive methods (beyond traditional ERP analyses) to obtain more holistic insights into error processing in complex choices. In our first study, using our novel complex choice task, we replicated common variations of error processing mechanisms with error detection, uncovered that a complex self-evaluation process is associated with changes in error processing mechanisms, and that self-evaluation has a broad and early influence on processing mechanisms. By modifying this complex task in our second study, we evoked motor inhibition errors and interference suppression errors and enabled a direct comparison between these two error types. For motor inhibition, we observed error-specific variations along the entire processing stream. For interference suppression, the complex response selection process renders the evaluation of conflict and inhibition particularly resource-intense independent of response accuracy. Our findings give raise to a variety of further research questions, underline the great potentials of combining different methodological approaches for future error processing research, and exemplify that the 8ART is a well-suited task to integrate different methodologies. Investigating the cognitive dynamics of error commission and error processing in complex choices by combining traditional ERP analyses with machine learning techniques, eye tracking parameters, computational modelling, and measures of peripheral motor execution in future work will expand our neurocognitive perspective on errors considerably.

References

Allain, S., Carbonnell, L., Burle, B., Hasbroucq, T., & Vidal, F. (2004). On-line executive control: An electromyographic study. *Psychophysiology*, *41*(1), 113–116. https://doi.org/10.1111/j.1469-8986.2003.00136.x

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bedoin, N., Abadie, R., Krzonowski, J., Ferragne, E., & Marcastel, A. (2019). A combined forced-attention dichotic listening – Go/Nogo task to assess response inhibition and interference suppression: An auditory event-related potential investigation. *Neuropsychology*, *33*(8), 1136–1150. https://doi.org/10.1037/neu0000586

Bekker, E. M., Kenemans, J. L., Hoeksma, M. R., Talsma, D., & Verbaten, M. N. (2005). The pure electrophysiology of stopping. *International Journal of Psychophysiology*, *55*(2), 191–198. https://doi.org/10.1016/j.ijpsycho.2004.07.005

Bender, A. D., Filmer, H. L., Garner, K. G., Naughtin, C. K., & Dux, P. E. (2016). On the relationship between response selection and response inhibition: An individual differences approach. *Attention, Perception, & Psychophysics*, *78*(8), 2420–2432. https://doi.org/10.3758/s13414-016-1158-8

Benikos, N., Johnstone, S. J., & Roodenrys, S. J. (2013). Varying task difficulty in the Go/Nogo task: The effects of inhibitory control, arousal, and perceived effort on ERP components. *International Journal of Psychophysiology*, *87*(3), 262–272. https://doi.org/10.1016/j.ijpsycho.2012.08.005

Bissett, P. G., Nee, D. E., & Jonides, J. (2009). Dissociating interference-control processes between memory and response. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1306–1316. https://doi.org/10.1037/a0016537

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014). State-of-the-Art of Visualization for Eye Tracking Data. *EuroVis - STARs*, 20 pages. https://doi.org/10.2312/EUROVISSTAR.20141173

Blasi, G., Goldberg, T. E., Weickert, T., Das, S., Kohn, P., Zoltick, B., Bertolino, A., Callicott, J. H., Weinberger, D. R., & Mattay, V. S. (2006). Brain regions underlying response inhibition and interference monitoring and suppression: Brain regions underlying cognitive control. *European Journal of Neuroscience*, *23*(6), 1658–1664. https://doi.org/10.1111/j.1460-9568.2006.04680.x

Bode, S., Feuerriegel, D., Bennett, D., & Alday, P. M. (2019). The Decision Decoding ToolBOX (DDTBOX) – A Multivariate Pattern Analysis Toolbox for Event-Related Potentials. *Neuroinformatics*, *17*(1), 27–42. https://doi.org/10.1007/s12021-018-9375-z

Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L., & Stahl, J. (2012). Predicting Perceptual Decision Biases from Early Brain Activity. *Journal of Neuroscience*, *32*(36), 12488–12498. https://doi.org/10.1523/JNEUROSCI.1708-12.2012

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology*, *103*, 357–369. https://doi.org/10.1016/j.biopsycho.2014.10.002

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology, 72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Bokura, H., Yamaguchi, S., & Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clinical Neurophysiology, 112*(12), 2224–2232. https://doi.org/10.1016/S1388-2457(01)00691-5

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing, 28*(2), 151–155. https://doi.org/10.1087/20150211

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153-178. 10.1016/j.cogpsych.2007.12.002

Bryce, D., Szűcs, D., Soltész, F., & Whitebread, D. (2011). The development of inhibitory control: An averaged and single-trial Lateralized Readiness Potential study. *NeuroImage, 57*(3), 671–685. https://doi.org/10.1016/j.neuroimage.2010.12.006

Brydges, C. R., Anderson, M., Reid, C. L., & Fox, A. M. (2013). Maturation of Cognitive Control: Delineating Response Inhibition and Interference Suppression. *PLoS ONE, 8*(7), e69826. https://doi.org/10.1371/journal.pone.0069826

Brydges, C. R., Clunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., Rousset, C., Whitelaw, P., Yeap, Y. J., & Fox, A. M. (2012). Dissociable Components of Cognitive Control: An Event-Related Potential (ERP) Study of Response Inhibition and Interference Suppression. *PLoS ONE, 7*(3), e34482. https://doi.org/10.1371/journal.pone.0034482

Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2020). The Role of Conflicting Representations and Uncertainty in Internal Error Detection During L2 Learning. *Language Learning*, *70*(S2), 75–103. https://doi.org/10.1111/lang.12401

Burgess, P. (1997). Theory and methodology in executive function research. In *Theory and Methodology of Frontal and Executive Function* (pp. 81–116). Psychology Press.

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error Negativity Does Not Reflect Conflict: A Reappraisal of Conflict Monitoring and Anterior Cingulate Cortex Activity. *Journal of Cognitive Neuroscience*, *20*(9), 1637–1655. https://doi.org/10.1162/jocn.2008.20110

Campbell, J. I. D., & Thompson, V. A. (2012). Morepower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Chang & Lin. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology,* 2:27:1-27:27.

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study: ERP measurement. *Psychophysiology*, *50*(2), 174–186. https://doi.org/10.1111/psyp.12001

Coles, M. G. H. (1989). Modern Mind-Brain Reading: Psychophysiology, Physiology, and Cognition. *Psychophysiology*, *26*(3), 251–269. https://doi.org/10.1111/j.1469-8986.1989.tb01916.x

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. https://doi.org/10.1016/S0301-0511(01)00076-X

Damaso, K., Williams, P., & Heathcote, A. (2020). Evidence for different types of errors being associated with different types of post-error changes. *Psychonomic Bulletin & Review*, *27*(3), 435–440. https://doi.org/10.3758/s13423-019-01675-w

Danielmeier, C., & Ullsperger, M. (2011). Post-Error Adjustments. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00233

Davidson, R. J., Jackson, D. C. (2000). Human electroencephalography. In *Handbook of psychophysiology* (pp. 27–52). US: Cambridge University Press.

de Jong, R., Coles, M. G. H., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 164–182. https://doi.org/10.1037/0096-1523.16.1.164

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a Neural System for Error Detection and Compensation. *Psychological Science*, *5*(5), 303–305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x

Dimoska, A., Johnstone, S. J., Barry, R. J., & Clarke, A. R. (2003). Inhibitory motor control in children with attention-deficit/hyperactivity disorder: Event-related potentials in the stop-signal paradigm. *Biological Psychiatry*, *54*(12), 1345–1354. https://doi.org/10.1016/S0006-3223(03)00703-0

Donchin, E. (1981). Surprise!? Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. https://doi.org/10.3758/s13414-011-0243-2

Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: Cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology*, *199*(3), 439–456. https://doi.org/10.1007/s00213-008-1127-6

Eimer, M. (1993). Spatial cueing, sensory gating and selective response preparation: An ERP study on visuo-spatial orienting. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(5), 408–420. https://doi.org/10.1016/0168-5597(93)90017-J

Endrass, T., Klawohn, J., Preuss, J., & Kathmann, N. (2012). Temporospatial dissociation of Pe subcomponents for perceived and unperceived errors. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00178

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: Aware and unaware errors in an antisaccade task: ERP correlates of conscious error recognition. *European Journal of Neuroscience*, *26*(6), 1714–1720. https://doi.org/10.1111/j.1460-9568.2007.05785.x

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Falkenstein, M. (2006). Inhibition, conflict and the Nogo-N2. *Clinical Neurophysiology*, *117*(8), 1638–1640. https://doi.org/10.1016/j.clinph.2006.05.002

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*(2–3), 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, *39*(1), 395–405. https://doi.org/10.1016/j.neuroimage.2007.09.001

Gehring, Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In *Oxford library of psychology. The Oxford handbook of event-related potential components* (pp. 231–291). Oxford University Press.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 198–216. https://doi.org/10.1037/0096-1523.18.1.198

Gehring, W. J., & Knight, R. T. (2000). Prefrontal–cingulate interactions in action monitoring. *Nature Neuroscience*, *3*(5), 516–520. https://doi.org/10.1038/74899

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gotzhein & Elson. (2020). *CORTEX-Computer-Aided Registration Tool for Experiments [online]*. University of Cologne. http://cortex. uni-koeln.de

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 331–344. https://doi.org/10.1037/0096-1523.14.3.331

Groom, M. J., & Cragg, L. (2015). Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. *Brain and Cognition*, *97*, 1–9. https://doi.org/10.1016/j.bandc.2015.04.004

Grützmann, R., Endrass, T., Klawohn, J., & Kathmann, N. (2014). Response accuracy rating modulates ERN and Pe amplitudes. *Biological Psychology*, *96*, 1–7. https://doi.org/10.1016/j.biopsycho.2013.10.007

Haider, M., Spong, P., & Lindsley, D. B. (1964). Attention, Vigilance, and Cortical Evoked-Potentials in Humans. *Science*, *145*(3628), 180–182. https://doi.org/10.1126/science.145.3628.180

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–197. https://doi.org/10.1016/j.bandc.2003.11.001

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*(2), 151–160. https://doi.org/10.1111/j.1469-8986.2005.00270.x

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189–202. https://doi.org/10.1016/j.actpsy.2010.04.011

Horváth, K., Kardos, Z., Takács, Á., Csépe, V., Nemeth, D., Janacsek, K., & Kóbor, A. (2020). Error Processing During the Online Retrieval of Probabilistic Sequence Knowledge. *Journal of Psychophysiology*, 1–15. https://doi.org/10.1027/0269-8803/a000262

Hsieh, S., Li, T.-H., & Tsai, L.-L. (2010). Impact of Monetary Incentives on Cognitive Performance and Error Monitoring following Sleep Deprivation. *Sleep*, *33*(4), 499–507. https://doi.org/10.1093/sleep/33.4.499

Hutchison, K. E., Niaura, R., & Swift, R. (1999). Smoking cues decrease prepulse inhibition of the startle response and increase subjective craving in humans. *Experimental and Clinical Psychopharmacology*, *7*(3), 250–256. https://doi.org/10.1037/1064-1297.7.3.250

Janczyk, M. (2017). A common capacity limitation for response and item selection in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1690–1698. https://doi.org/10.1037/xlm0000408

Jasper. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, *10*(2), 370–375. https://doi.org/10.1016/0013-4694(58)90053-1

Jepma, M., & Nieuwenhuis, S. (2011). Pupil Diameter Predicts Changes in the Exploration–Exploitation Trade-off: Evidence for the Adaptive Gain Theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596. https://doi.org/10.1162/jocn.2010.21548

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*(1), 90–97. https://doi.org/10.1111/j.1469-8986.1993.tb03208.x

Kaczkurkin, A. N. (2013). The effect of manipulating task difficulty on error-related negativity in individuals with obsessive-compulsive symptoms. *Biological Psychology*, *93*(1), 122–131. https://doi.org/10.1016/j.biopsycho.2013.01.001

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed). Prentice-Hall.

Kirschner, H., Humann, J., Derrfuss, J., Danielmeier, C., & Ullsperger, M. (2021). Neural and behavioral traces of error awareness. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 573–591. https://doi.org/10.3758/s13415-020-00838-w

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., & Ullsperger, M. (2007). Genetically Determined Differences in Learning from Errors. *Science*, *318*(5856), 1642–1645. https://doi.org/10.1126/science.1145044

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. https://doi.org/10.1017/S0048577201990559

Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P. H., & Ridderinkhof, K. R. (2004). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology*, *41*(1), 9–20. https://doi.org/10.1046/j.1469-8986.2003.00127.x

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*(3), 282–294. https://doi.org/10.1111/j.1469-8986.1996.tb00425.x

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., Gerard Moeller, F., & Swann, A. C. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology*, *46*(5), 1059–1068. https://doi.org/10.1111/j.1469-8986.2009.00845.x

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In *Inhibitory processes in attention, memory, and language* (pp. 189–239). Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327. https://doi.org/10.1037/0033-295X.91.3.295

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition). The MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and Clinical Neurophysiology*, *75*(6), 528–542. https://doi.org/10.1016/0013-4694(90)90139-B

Maier, M. E., Ernst, B., & Steinhauser, M. (2019). Error-related pupil dilation is sensitive to the evaluation of different error types. *Biological Psychology*, *141*, 25–34. https://doi.org/10.1016/j.biopsycho.2018.12.013

Maier, M., Steinhauser, M., & Hübner, R. (2010). Effects of response-set size on error-related brain activity. *Experimental Brain Research*, *202*(3), 571–581. https://doi.org/10.1007/s00221-010-2160-3

Maier, Pellegrino, G., & Steinhauser, M. (2012). Enhanced error-related negativity on flanker errors: Error expectancy or error significance?: Ne/ERN and flanker errors. *Psychophysiology*, *49*(7), 899–908. https://doi.org/10.1111/j.1469-8986.2012.01373.x

Maier, & Steinhauser. (2016). Error significance but not error expectancy predicts error-related negativities for different error types. *Behavioural Brain Research*, *297*, 259–267. https://doi.org/10.1016/j.bbr.2015.10.031

Maier, Steinhauser, & Hübner, R. (2008). Is the Error-related Negativity Amplitude Related to Error Detectability? Evidence from Effects of Different Error Types. *Journal of Cognitive Neuroscience*, *20*(12), 2263–2273. https://doi.org/10.1162/jocn.2008.20159

Maier, Yeung, N., & Steinhauser, M. (2011). Error-related brain activity and adjustments of selective attention following errors. *NeuroImage*, *56*(4), 2339–2347. https://doi.org/10.1016/j.neuroimage.2011.03.083

Mangun, G. A., Hillyard, S. A., & Luck, S. J. (1993). Electrocortical substrates of visual selective attention. In *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 219–243). The MIT Press.

Maruo, Y. (2017). The effect of monetary punishment on error evaluation in a Go/No-go task. *International Journal of Psychophysiology*, 6.

Maruo, Y., Schacht, A., Sommer, W., & Masaki, H. (2016). Impacts of motivational valence on the error-related negativity elicited by full and partial errors. *Biological Psychology*, *114*, 108–116. https://doi.org/10.1016/j.biopsycho.2015.12.004

Maruo, Y., Sommer, W., & Masaki, H. (2017). The effect of monetary punishment on error evaluation in a Go/No-go task. *International Journal of Psychophysiology*, *120*, 54–59. https://doi.org/10.1016/j.ijpsycho.2017.07.002

Mattes, A., Porth, E., & Stahl, J. (2022). Linking neurophysiological processes of action monitoring to post-response speed-accuracy adjustments in a neuro-cognitive diffusion model. *NeuroImage*, *247*, 118798. https://doi.org/10.1016/j.neuroimage.2021.118798

Medina, J. M., Wong, W., DÃ-az, J. A., & Colonius, H. (2015). Advances in modern mental chronometry. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00256

Meyer, A., Riesel, A., & Proudfit, G. H. (2013). Reliability of the ERN across multiple tasks as a function of increasing errors: Reliability of the ERN across multiple tasks. *Psychophysiology*, *50*(12), 1220–1225. https://doi.org/10.1111/psyp.12132

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, *26*(1–3), 3–67. https://doi.org/10.1016/0301-0511(88)90013-0

Miller, J., & Hackley, S. A. (1992). Electrophysiological evidence for temporal overlap among contingent mental processes. *Journal of Experimental Psychology: General*, *121*(2), 195–209. https://doi.org/10.1037/0096-3445.121.2.195

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*(1), 99–115. https://doi.org/10.1111/1469-8986.3510099

Moser, J. S., Moran, T. P., Schroder, H. S., Donnellan, M. B., & Yeung, N. (2013). On the relationship between anxiety and error monitoring: A meta-analysis and conceptual framework. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00466

Mück, M., Mattes, A., Porth, E., & Stahl, J. (2023). Narcissism and the perception of failure - evidence from the error-related negativity and the error positivity. *Personality Neuroscience*, *6*, e2. https://doi.org/10.1017/pen.2022.7

Mück, M., Ohmann, K., Dummel, S., Mattes, A., Thesing, U., & Stahl, J. (2020). Face Perception and Narcissism: Variations of Event-Related Potential Components (P1 & N170) with Admiration and Rivalry. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(5), 1041–1055. https://doi.org/10.3758/s13415-020-00818-0

Murphy, P. R., van Moort, M. L., & Nieuwenhuis, S. (2016). The Pupillary Orienting Response Predicts Adaptive Behavioral Adjustment after Errors. *PLOS ONE, 11*(3), e0151763. https://doi.org/10.1371/journal.pone.0151763

Nayak, S., Salem, H. Z., & Tarullo, A. R. (2020). Neural mechanisms of response-preparation and inhibition in bilingual and monolingual children: Lateralized Readiness Potentials (LRPs) during a nonverbal Stroop task. *Developmental Cognitive Neuroscience, 41*, 100740. https://doi.org/10.1016/j.dcn.2019.100740

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(1), 1–17. https://doi.org/10.3758/CABN.7.1.1

Niessen, E., Ant, J. M., Bode, S., Saliger, J., Karbe, H., Fink, G. R., Stahl, J., & Weiss, P. H. (2020). Preserved performance monitoring and error detection in left hemisphere stroke. *NeuroImage: Clinical*, *27*, 102307. https://doi.org/10.1016/j.nicl.2020.102307

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246. https://doi.org/10.1037/0033-2909.126.2.220

Novikov, N. A., Nurislamova, Y. M., Zhozhikashvili, N. A., Kalenkovich, E. E., Lapina, A. A., & Chernyshev, B. V. (2017). Slow and Fast Responses: Two Mechanisms of Trial Outcome Processing Revealed by EEG Oscillations. *Frontiers in Human Neuroscience*, *11*, 218. https://doi.org/10.3389/fnhum.2017.00218

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., & Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: A high-density electrical mapping study: Error awareness. *European Journal of Neuroscience*, *25*(8), 2571–2579. https://doi.org/10.1111/j.1460-9568.2007.05477.x

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Turin, A., Ward, S., Foxe, J. J., & Robertson, I. H. (2009). Two Types of Action Error: Electrophysiological Evidence for Separable Inhibitory and Sustained Attention Neural Mechanisms Producing Error on Go/No-go Tasks. *Journal of Cognitive Neuroscience*, *21*(1), 93–104. https://doi.org/10.1162/jocn.2009.21008

O'Keeffe, F. M., Dockree, P. M., Moloney, P., Carton, S., & Robertson, I. H. (2007). Characterising error-awareness of attentional lapses and inhibitory control failures in patients with traumatic brain injury. *Experimental Brain Research*, *180*(1), 59–67. https://doi.org/10.1007/s00221-006-0832-9

Olvet, D. M., & Hajcak, G. (2009a). The stability of error-related brain activity with increasing trials. *Psychophysiology, 46*(5), 957–961. https://doi.org/10.1111/j.1469-8986.2009.00848.x

Olvet, D. M., & Hajcak, G. (2009b). The effect of trial-to-trial feedback on the error-related negativity and its relationship with anxiety. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 427–433. https://doi.org/10.3758/CABN.9.4.427

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable Components of Error Processing: On the Functional Significance of the Pe Vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Overhoff, H., Ko, Y. H., Fink, G. R., Stahl, J., Weiss, P. H., Bode, S., & Niessen, E. (2022). The relationship between response dynamics and the formation of confidence varies across the lifespan. *Frontiers in Aging Neuroscience*, *14*, 969074. https://doi.org/10.3389/fnagi.2022.969074

Pailing, P. E., & Segalowitz, S. J. (2004). The effects of uncertainty in error monitoring on associated ERPs. *Brain and Cognition*, *56*(2), 215–233. https://doi.org/10.1016/j.bandc.2004.06.005

Paul, K., Walentowska, W., Bakic, J., Dondaine, T., & Pourtois, G. (2017). Modulatory effects of happy mood on performance monitoring: Insights from error-related brain potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(1), 106–123. https://doi.org/10.3758/s13415-016-0466-8

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Pfister, R., & Foerster, A. (2022). How to measure post-error slowing: The case of pre-error speeding. *Behavior Research Methods*, *54*(1), 435–443. https://doi.org/10.3758/s13428-021-01631-4

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Porth, E., Mattes, A., & Stahl, J. (2022). The influence of error detection and error significance on neural and behavioral correlates of error processing in a complex choice task. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(6), 1231–1249. https://doi.org/10.3758/s13415-022-01028-6

Porth, E., Mattes, A., & Stahl, J. (2023). *Motor Inhibition versus Interference Suppression: Neural and Behavioural Features of Action Monitoring and Error Processing* [Manuscript submitted for publication]. Department Psychology, University of Cologne.

Posner, M. I. (1978). *Chronometric explorations of mind*. Lawrence Erlbaum Associates.

Posner, M. I. (2005). Timing the Brain: Mental Chronometry as a Tool in Neuroscience. *PLoS Biology*, *3*(2), e51. https://doi.org/10.1371/journal.pbio.0030051

Proverbio, A. M., Cozzi, M., Orlandi, A., & Carminati, M. (2017). Error-related negativity in the skilled brain of pianists reveals motor simulation. *Neuroscience*, *346*, 309–319. https://doi.org/10.1016/j.neuroscience.2017.01.030

Rabbitt, P. M. A. (1968). Three Kinds of Error-Signalling Responses in a Serial Choice Task. *Quarterly Journal of Experimental Psychology*, *20*(2), 179–188. https://doi.org/10.1080/14640746808400146

Rabbitt, & Rodgers, B. (1977). What does a Man do after he Makes an Error? An Analysis of Response Programming. *Quarterly Journal of Experimental Psychology*, *29*(4), 727–743. https://doi.org/10.1080/14640747708400645

Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: The N2/P3 complex further validated. *Brain and Cognition*, *56*(2), 234–252. https://doi.org/10.1016/j.bandc.2004.07.002

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review*, *95*(2), 238–255. https://doi.org/10.1037/0033-295X.95.2.238

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

Ridderinkhof, R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, *66*(4), 312–323. https://doi.org/10.1007/s00426-002-0104-7

Riesel, A., Härpfer, K., Thoma, L., Kathmann, N., & Klawohn, J. (2023). Associations of neural error-processing with symptoms and traits in a dimensional sample recruited across the obsessive-compulsive spectrum. *Psychophysiology*, *60*(2), e14164. https://doi.org/10.1111/psyp.14164

Riesel, A., Kathmann, N., Wüllhorst, V., Banica, I., & Weinberg, A. (2019). Punishment has a persistent effect on error-related brain activity in highly anxious individuals twenty-four hours after conditioning. *International Journal of Psychophysiology*, *146*, 63–72. https://doi.org/10.1016/j.ijpsycho.2019.09.014

Riesel, A., Richter, A., Kaufmann, C., Kathmann, N., & Endrass, T. (2015). Performance monitoring in obsessive–compulsive undergraduates: Effects of task difficulty. *Brain and Cognition*, *98*, 35–42. https://doi.org/10.1016/j.bandc.2015.05.002

Riesel, A., Weinberg, A., Endrass, T., Kathmann, N., & Hajcak, G. (2012). Punishment has a lasting impact on error-related brain activity: Punishment modulates error monitoring. *Psychophysiology*, *49*(2), 239–247. https://doi.org/10.1111/j.1469-8986.2011.01298.x

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*, *93*(3), 377–385. https://doi.org/10.1016/j.biopsycho.2013.04.007

Rietdijk, W. J. R., Franken, I. H. A., & Thurik, A. R. (2014). Internal Consistency of Event-Related Potentials Associated with Cognitive Control: N2/P3 and ERN/Pe. *PLoS ONE*, *9*(7), e102672. https://doi.org/10.1371/journal.pone.0102672

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rondeel, E. W. M., van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00494

Ruiz, M. H., Jabusch, H.-C., & Altenmüller, E. (2009). Detecting Wrong Notes in Advance: Neuronal Correlates of Error Monitoring in Pianists. *Cerebral Cortex*, *19*(11), 2625–2639. https://doi.org/10.1093/cercor/bhp021

Sayette, M. A., Griffin, K. M., & Sayers, W. M. (2010). Counterbalancing in Smoking Cue Research: A Critical Analysis. *Nicotine & Tobacco Research*, *12*(11), 1068–1079. https://doi.org/10.1093/ntr/ntq159

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141

Schreiber, M., Endrass, T., Weigand, A., & Kathmann, N. (2012). Age Effects on Adjustments of Performance Monitoring to Task Difficulty. *Journal of Psychophysiology*, *26*(4), 145–153. https://doi.org/10.1027/0269-8803/a000077

Sebastian, A., Gerdes, B., Feige, B., Klöppel, S., Lange, T., Philipsen, A., Tebartz van Elst, L., Lieb, K., & Tüscher, O. (2012). Neural correlates of interference inhibition, action withholding and action cancelation in adult ADHD. *Psychiatry Research: Neuroimaging*, *202*(2), 132–141. https://doi.org/10.1016/j.pscychresns.2012.02.010

Selimbeyoglu, A., Keskin-Ergen, Y., & Demiralp, T. (2012). What if you are not sure? Electroencephalographic correlates of subjective confidence level about a decision. *Clinical Neurophysiology*, *123*(6), 1158–1167. https://doi.org/10.1016/j.clinph.2011.10.037

Shalgi, S., O'Connell, R. G., Deouell, L. Y., & Robertson, I. H. (2007). Absent minded but accurate: Delaying responses increases accuracy but decreases error awareness. *Experimental Brain Research*, *182*(1), 119–124. https://doi.org/10.1007/s00221-007-1054-5

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Siswandari, Y., Bode, S., & Stahl, J. (2019). Performance monitoring beyond choice tasks: The time course of force execution monitoring investigated by event-related potentials and multivariate pattern analysis. *NeuroImage*, *197*, 544–556. https://doi.org/10.1016/j.neuroimage.2019.05.006

Smulders, F. T. Y. (2010). Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies. *Psychophysiology*, *47*(2), 387–392. https://doi.org/10.1111/j.1469-8986.2009.00934.x

Smulders, F. T. Y., & Miller, J. O. (2011). *The Lateralized Readiness Potential*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0115

Somon, B., Campagne, A., Delorme, A., & Berberian, B. (2019). Evaluation of performance monitoring ERPs through difficulty manipulation in a response-feedback paradigm. *Brain Research*, *1704*, 196–206. https://doi.org/10.1016/j.brainres.2018.10.007

Spunt, R. P., Lieberman, M. D., Cohen, J. R., & Eisenberger, N. I. (2012). The Phenomenology of Error Processing: The Dorsal ACC Response to Stop-signal Errors Tracks Reports of Negative Affect. *Journal of Cognitive Neuroscience*, *24*(8), 1753–1765. https://doi.org/10.1162/jocn_a_00242

Stahl, J., Acharki, M., Kresimon, M., Völler, F., & Gibbons, H. (2015). Perfect error processing: Perfectionism-related variations in action monitoring and error processing mechanisms. *International Journal of Psychophysiology*, *97*(2), 153–162. https://doi.org/10.1016/j.ijpsycho.2015.06.002

Stahl, J., & Gibbons, H. (2004). The application of jackknife-based onset detection of lateralized readiness potential in correlative approaches. *Psychophysiology*, *41*(6), 845–860. https://doi.org/10.1111/j.1469-8986.2004.00243.x

Stahl, J., Mattes, A., Hundrieser, M., Kummer, K., Mück, M., Niessen, E., Porth, E., Siswandari, Y., Wolters, P., & Dummel, S. (2020). Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task. *Biological Psychology*, *156*, 107969. https://doi.org/10.1016/j.biopsycho.2020.107969

Stahl, J., & Rammsayer, T. (2004). Differences in the transmission of sensory input into motor output between introverts and extraverts: Behavioral and psychophysiological analyses. *Brain and Cognition*, *56*(3), 293–303. https://doi.org/10.1016/j.bandc.2004.07.004

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, *52*(1), 77–86. https://doi.org/10.1016/j.ijpsycho.2003.12.005

Steinhauser, M., & Yeung, N. (2010). Decision Processes in Human Performance Monitoring. *Journal of Neuroscience*, *30*(46), 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010

Steinhauser, M., & Yeung, N. (2012). Error awareness as evidence accumulation: Effects of speed-accuracy trade-off on error signaling. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00240

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tops, M., & Boksem, M. A. S. (2010). Absorbed in the task: Personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(4), 441–453. https://doi.org/10.3758/CABN.10.4.441

Tops, M., Boksem, M. A. S., Wester, A. E., Lorist, M. M., & Meijman, T. F. (2006). Task engagement and the relationships between the error-related negativity, agreeableness, behavioral shame proneness and cortisol. *Psychoneuroendocrinology*, *31*(7), 847–858. https://doi.org/10.1016/j.psyneuen.2006.04.001

Ullsperger, M., & von Cramon, D. Y. (2004). Neuroimaging of Performance Monitoring: Error Detection and Beyond. *Cortex*, *40*(4–5), 593–604. https://doi.org/10.1016/S0010-9452(08)70155-2

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*(5), 816–827. https://doi.org/10.1111/1469-8986.3850816

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

van Driel, J., Ridderinkhof, K. R., & Cohen, M. X. (2012). Not All Errors Are Alike: Theta and Alpha EEG Dynamics Relate to Differences in Error-Processing Dynamics. *Journal of Neuroscience*, *32*(47), 16795–16806. https://doi.org/10.1523/JNEUROSCI.0802-12.2012

Van Veen, V. van, & Carter, C. S. (2002). The Timing of Action-Monitoring Processes in the Anterior Cingulate Cortex. *Journal of Cognitive Neuroscience*, *14*(4), 593–602. https://doi.org/10.1162/08989290260045837

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity' specific to errors? *Biological Psychology*, *51*(2–3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0

Voodla, A., & Uusberg, A. (2021). Do performance-monitoring related cortical potentials mediate fluency and difficulty effects on decision confidence? *Neuropsychologia*, *155*, 107822. https://doi.org/10.1016/j.neuropsychologia.2021.107822

Vuillier, L., Bryce, D., Szücs, D., & Whitebread, D. (2016). The Maturation of Interference Suppression and Response Inhibition: ERP Analysis of a Cued Go/Nogo Task. *PLOS ONE, 11*(11), e0165697. https://doi.org/10.1371/journal.pone.0165697

Wessel, J. R. (2012). Error awareness and the error-related negativity: Evaluating the first decade of evidence. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00088

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, *55*(3), e13041. https://doi.org/10.1111/psyp.13041

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error Awareness Revisited: Accumulation of Multimodal Evidence from Central and Autonomic Nervous Systems. *Journal of Cognitive Neuroscience*, *23*(10), 3021–3036. https://doi.org/10.1162/jocn.2011.21635

Winer, B. J. (1971). *Statistical principles in experimental design* (2d ed). McGraw-Hill.

Wiswede, D., Rothermund, K., & Frings, C. (2013). Not all errors are created equally: Specific ERN responses for errors originating from distractor-based response retrieval. *European Journal of Neuroscience, 38*(10), 3496–3506. https://doi.org/10.1111/ejn.12340

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Supplementary Materials of Study 1

Comparison of Excluded and Included Subsamples

For RT, there was no significant effect of sample, $F(1, 36) = 1.41$, $p = .244$, $\eta_p^2 = .04$, and there was no significant interaction of response type and sample, $F(1.21, 43.47) = 0.95$, $p = .351$, $\eta_p^2 = .03$. For peak RF, there was no significant effect for sample, $F(1, 36) = 2.09$, $p = .157$, $\eta_p^2 = .05$, and no significant interaction of response type and sample, $F(1.40, 50.47) = 0.95$, $p = .364$, $\eta_p^2 = .03$. For the pre-post RT_{diff}, neither sample, $F(1, 35) = 1.65$, $p = .207$, $\eta_p^2 = .05$, nor the interaction of response type and sample, $F(1.21, 42.38) = 0.39$, $p = .578$, $\eta_p^2 = .01$, yielded significant effects.

Electrophysiological Results of the Untransformed Event-related Potentials

The descriptive statistics of the untransformed ERP data of the three response types are depicted in Table S1-1.

Table S1-1.

Means and standard error of means of the untransformed data for the $N_{e/c}$ amplitude, the $P_{e/c}$ amplitude, and the $P_{e/c}$ activity separately for each response type for n = 21.

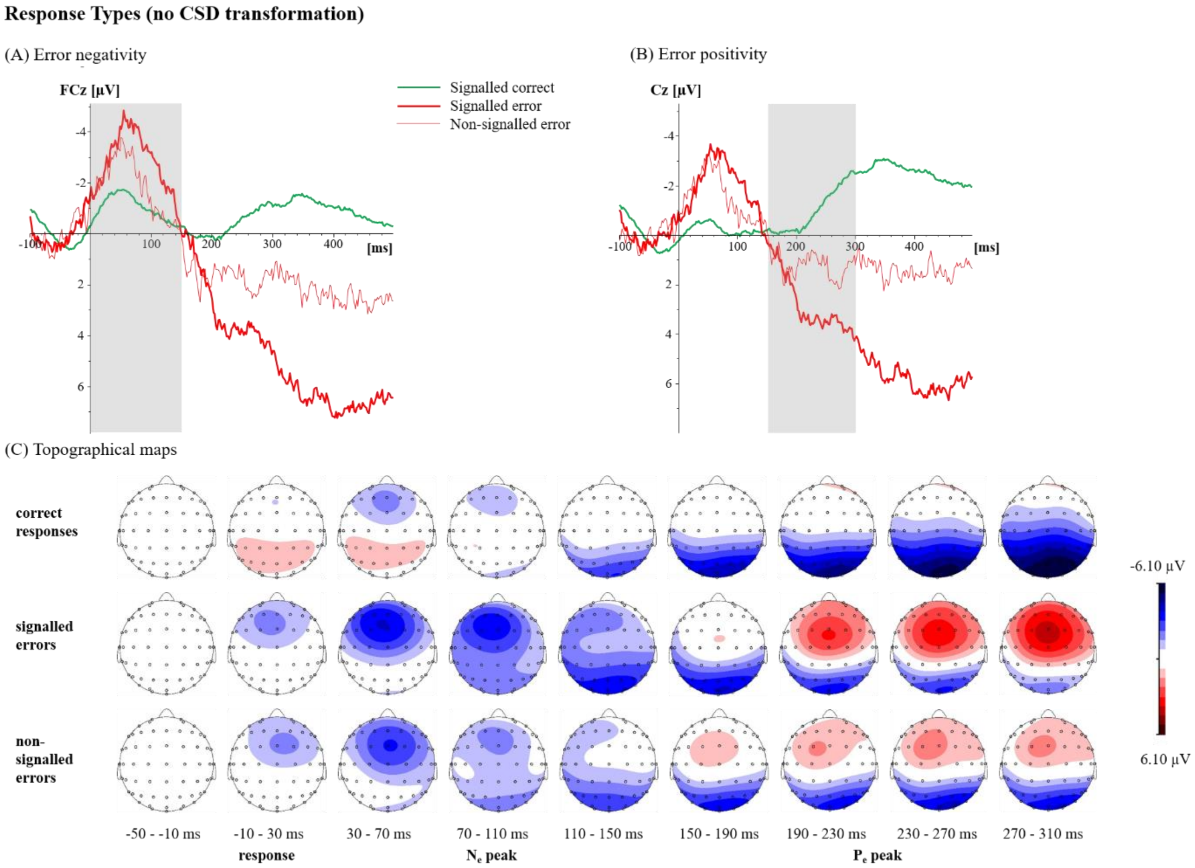
	Signalled Correct		**Signalled Errors**		**Non-signalled Errors**	
	M	**SE**	**M**	**SE**	**M**	**SE**
$N_{e/c}$ amplitude [µV]	-2.1	0.5	-5.8	0.8	-5.3	0.7
$P_{e/c}$ amplitude [µV]	0.6	0.7	5.6	1.0	4.3	0.9
$P_{e/c}$ mean activity [µV]	0.3	0.6	4.0	1.1	2.1	0.8

The pattern of results for the $N_{e/c}$ amplitude was similar to the CSD transformed data. The $N_{e/c}$ amplitude varied significantly with response types, $F(1.39, 27.79) = 11.26$, $p < .001$, $\eta_p^2 = .36$. The $N_{e/c}$ amplitude was larger for signalled errors and non-signalled errors compared to signalled correct responses, $p < .001$ and $p = .002$, respectively, and the two error types did not differ significantly, $p = .799$. For the $P_{e/c}$ the pattern of results slightly differed from the CSD analyses. The $P_{e/c}$ amplitude also varied with response type,

$F(2, 40) = 12.93$, $p < .001$, $\eta_p^2 = .39$, and was larger for signalled and non-signalled errors compared to signalled correct responses, $p < .001$ and $p = .002$, respectively. Opposed to the CSD transformed data, the two error types did not differ significantly in $P_{e/c}$ amplitude, $p = .415$. Analogously to the CSD transformed data, the $P_{e/c}$ mean activity varied significantly with response type, $F(2, 40) = 6.67$, $p = .003$, $\eta_p^2 = .25$, and was larger for signalled errors compared to signalled correct responses, $p = .002$, while it did not differ significantly between non-signalled errors and correct responses, $p = .197$. The difference between signalled and non-signalled errors was not significant, $p = .154$. The untransformed grand average waveforms are depicted in Figure S1-1.

Figure S1-1

Averaged untransformed ERP waveforms for the three response types



Note. Averaged untransformed waveforms of the ERP component (A) error negativity, measured at FCz electrode as the mean amplitude (± 2 data points) around the negative peak in the time window of 0-150 ms after response onset and (B) error positivity, measured at Cz electrode as the mean amplitude (± 2 data points) around the positive peak and as mean activity in the time window of 150-300 ms after response onset grouped by response type, as well as the respective topographical maps for errors of both conditions.

The descriptive statistics of the untransformed ERP data from the no-self-evaluation condition and the self-evaluation condition are depicted in Table S1-2.

Table S1-2

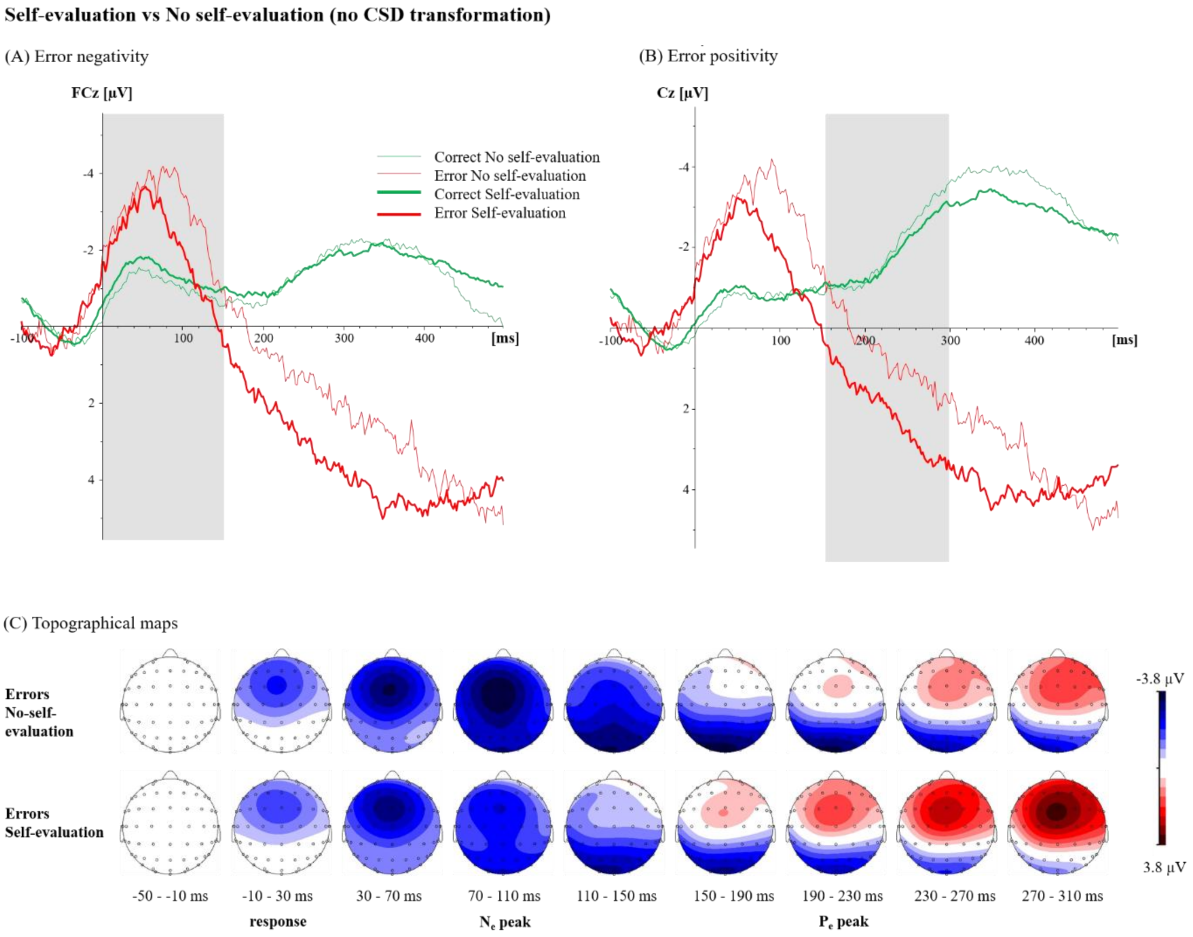
Means and standard error of means of the untransformed data for the $N_{e/c}$ amplitude, the $P_{e/c}$ amplitude, and the $P_{e/c}$ activity for correct and erroneous responses of the no-self-evaluation and the self-evaluation condition.

	No self-evaluation				**Self-evaluation**			
	Correct		**Error**		**Correct**		**Error**	
	M	**SE**	**M**	**SE**	**M**	**SE**	**M**	**SE**
$N_{e/c}$ amplitude [µV]	-2.3	0.4	-5.7	0.5	-2.4	0.4	-4.4	0.4
$P_{e/c}$ amplitude [µV]	-0.3	0.5	3.5	0.7	-0.4	0.5	4.2	0.6
$P_{e/c}$ mean activity [µV]	-0.8	0.6	2.1	0.7	-0.7	0.5	3.2	0.6

For the $N_{e/c}$ amplitude the analyses of the untransformed ERPs showed a significant main effect of accuracy, $F(1, 32) = 54.07$, $p < .001$, $\eta_p^2 = .63$, and of self-evaluation, $F(1, 32) = 7.00$, $p = .013$, $\eta_p^2 = .18$, and a significant interaction of accuracy and self-evaluation, $F(1, 32) = 11.56$, $p = .002$, $\eta_p^2 = .27$. The N_e amplitude was larger for errors than for correct responses in the no-self-evaluation condition and in the self-evaluation condition, both p-values < .001. The $N_{e/c}$ amplitude for errors was smaller in the self-evaluation condition than in the no-self-evaluation condition, $p < .001$, while it did not differ significantly for correct responses, $p = .989$. Similar to the CSD-transformed data we only observed a significant effect of Accuracy for the $P_{e/c}$ amplitude, $F(1, 32) = 37.39$, $p < .001$, $\eta_p^2 = .54$, while there was no significant effect of self-evaluation, $F(1, 32) = 0.87$, $p = .357$, $\eta_p^2 = .03$, and no significant interaction of accuracy and self-evaluation, $F(1, 32) = 2.36$, $p = .135$, $\eta_p^2 = .07$. For the $P_{e/c}$ mean activity there was a significant effect of accuracy, $F(1, 32) = 26.54$, $p < .001$, $\eta_p^2 = .45$, but no significant main effect of self-evaluation, $F(1, 32) = 3.37$, $p = .076$, $\eta_p^2 = .10$. The interaction of accuracy and self-evaluation did not reach significance, $F(1, 32) = 2.82$, $p = .103$, $\eta_p^2 = .08$. The untransformed grand average waveforms are depicted in Figure S1-2.

Figure S1-2

Averaged untransformed ERP waveforms for the two self-evaluation conditions



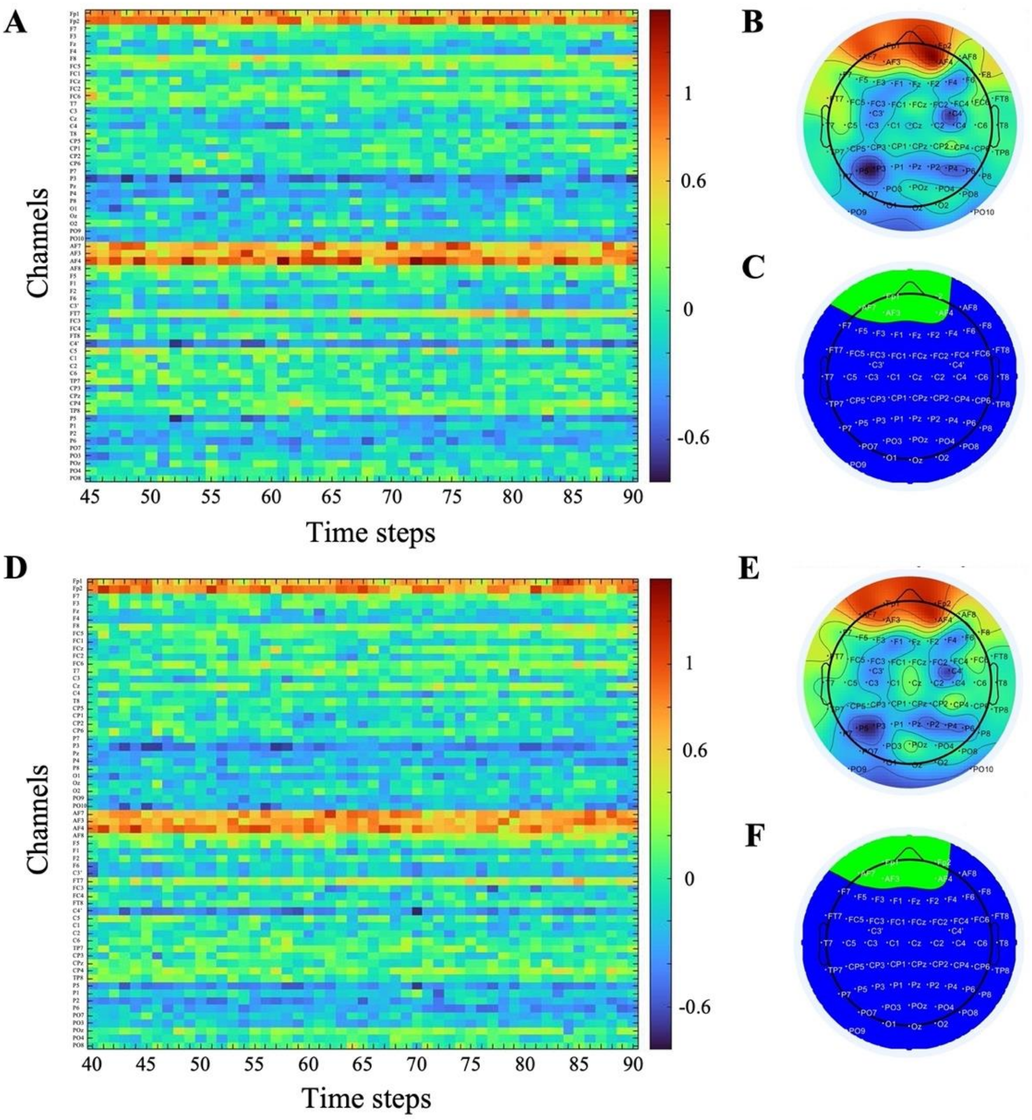
Note. Averaged *un*transformed waveforms of the ERP component (A) error negativity, measured at FCz electrode as the mean amplitude (± 2 data points) around the negative peak in the time window of 0-150 ms after response onset and (B) error positivity, measured at Cz electrode as the mean amplitude (± 2 data points) around the positive peak and as mean activity in the time window of 150-300 ms after response onset grouped by accuracy (correct, error) and experimental condition (no-self-evaluation, self-evaluation), as well as the respective topographical maps for errors of both conditions.

Feature Weights Analyses

We conducted feature weights analyses to investigate from which channels the classifier uses information to decode the conditions from brain activity. We computed feature weights for the contrast between correct responses of the no-self-evaluation condition and correct responses of the self-evaluation condition, as well as for the contrast between errors from both conditions, as for these two contrasts the classifier was successful in decoding the condition from brain activity. We computed the feature weights for the time steps where the classifier decoded the condition significantly above chance (time steps 45 to 90 for correct responses and time steps 40 to 90 for errors). First, for each channel (i.e. each feature) the absolute feature weights were averaged within each of the 10 ms time steps. The resulting values were z-standardized for each time step. The significance of the feature weights of each channel and each time step was tested with a series of Bonferroni-corrected *t*-tests. A significant feature weight indicates that the channel substantially contributes to the decoding accuracy of the classifier. The results of the feature weights analyses are depicted in Figure S1-3.

Figure S1-3

Results from the feature weights analyses

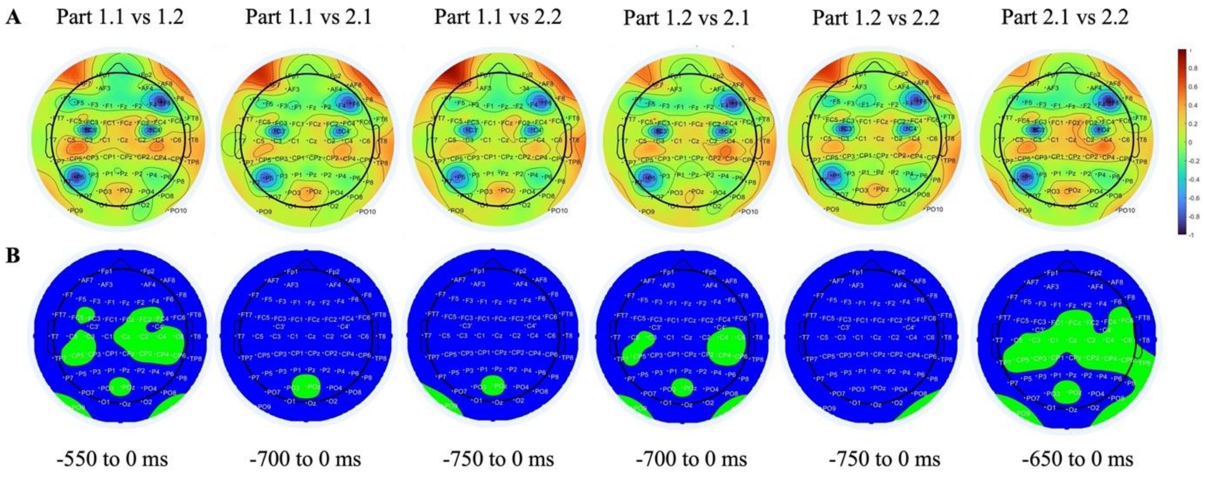


Note. Z-standardized absolute feature weights averaged across the significant pre-response time steps (A and D) with the respective topographical maps (B and E) and the statistical threshold maps where significant features are color-coded green (C and F) for the contrast of correct responses of the no-self-evaluation condition vs correct responses of the self-evaluation condition (A-C) and for the contrast of errors of the no-self-evaluation condition vs errors of the self-evaluation condition (D-F).

The z-standardized absolute feature weights averaged across time steps show similar patterns for the contrast of the no-self-evaluation and the self-evaluation condition for both correct responses and errors. This suggests that the decoding success is not based on error specific processes. Significant features comprise anterior-frontal and fronto-parietal electrode sites. This is in line with the assumption that the MVPA results capture differences in broader cognitive control processes and executive functions between the two self-evaluation conditions. However, this interpretation has to be made with care, as the distribution of the signal on the scalp is only an approximation for the underlying neural generators. Moreover, significant features can also be important because they systematically bind irrelevant variance, thus suppressing noise for other features that supply meaningful information (Haufe et al., 2014). We also computed feature weights for the partwise MVPA contrasts. The results are depicted in Figure S1-4. Again, the results suggest that for all contrasts similar channels are important.

Figure S1-4

Results from the partwise feature weights analyses



Note. Z-standardized absolute feature weights averaged across the significant pre-response time steps mapped onto the scalp (A) and the respective statistical threshold maps where significant features are color-coded green (B) for the partwise contrasts of responses of the no-self-evaluation condition vs responses of the self-evaluation condition.

Trial-matched Event-related Potential Results

The no-self-evaluation condition comprises less trials (320 trials) than the self-evaluation condition (704 trials). To ensure that the measurement of the N_e in the shorter no-self-evaluation condition is as stable as in the longer self-evaluation condition, we trial-matched the four conditions (errors and correct responses from the no-self-evaluation condition and the self-evaluation condition) based on the condition with the smallest number of trials (separately for each participant) and randomly drew the same number of trials from the other three conditions. We computed our ERP analyses for this trial-matched sample. The descriptive statistics are depicted in Table S1-3.

Table S1-3

Means and standard error of means of the N_e amplitude (no-self-evaluation vs self-evaluation) for correct and erroneous responses for the non-trial-matched and the trial-matched data.

	No Self-Evaluation				**Self-Evaluation**			
	Correct		**Error**		**Correct**		**Error**	
	M	**SE**	**M**	**SE**	**M**	**SE**	**M**	**SE**
Non-trial-matched $N_{e/c}$ amplitude [µV/cm²]	-0.12	0.02	-0.20	0.03	-0.13	0.02	-0.16	0.02
Trial-matched $N_{e/c}$ amplitude [µV/cm²]	-0.15	0.02	-0.20	0.03	-0.18	0.03	-0.17	0.03

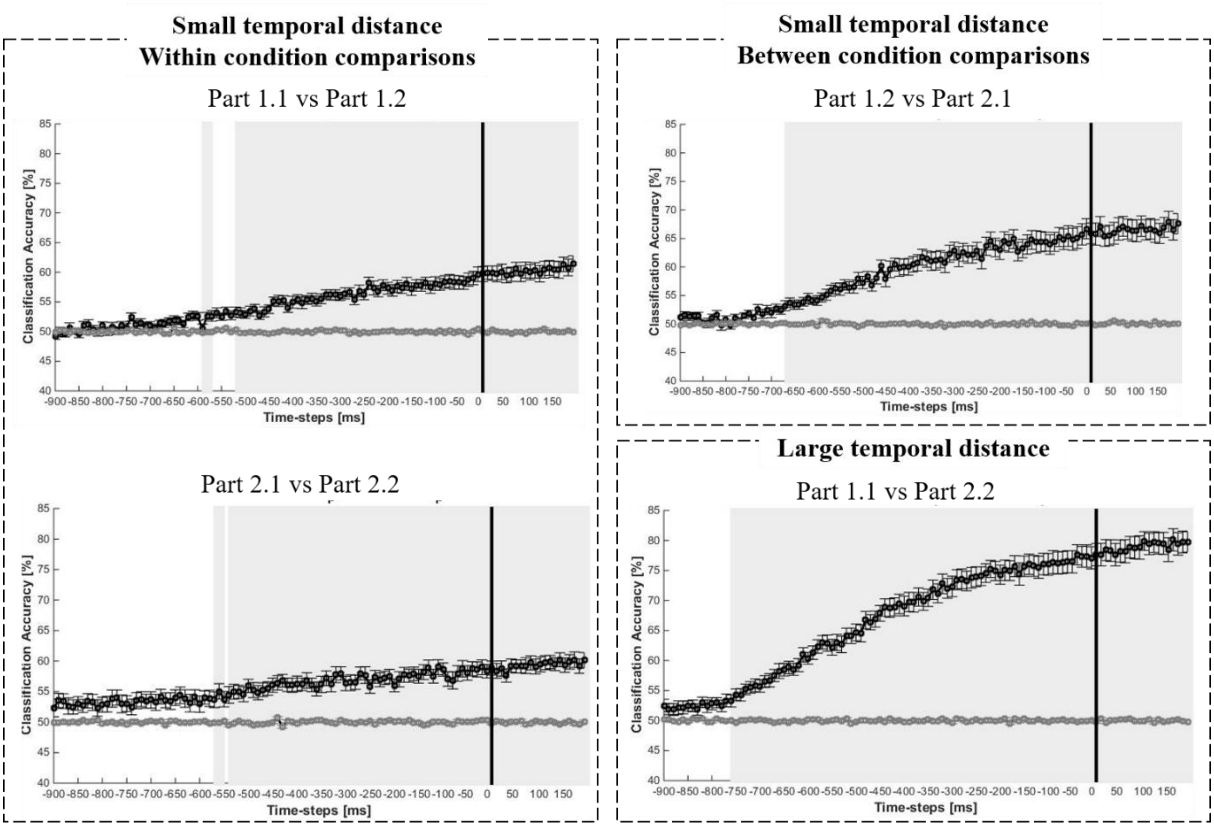
The trial-matched N_e amplitude and the non-trial-matched N_e amplitude did not differ significantly for errors in the no-self-evaluation condition, $t(32) = 1.84$, $p = .075$, $d = 0.32$, and for errors in the self-evaluation condition, $t(32) = 0.95$, $p = .349$, $d = 0.17$. We can thus rule out significant differences in N_e amplitude between the two conditions that are based on the number of trials.

Partwise Multivariate Pattern Analyses Matched for Response Times

To investigate whether or not the results of the partwise MVPA are influenced by the change in response times across the experiment, we conducted RT-matched partwise MVPA. In RT-matched analyses, for the first-level MVPA only trials with similar response times are included in the contrasts. The results from the RT-matched partwise MVPA are depicted in Figure S1-5.

Figure S1-5

RT-matched partwise MVPA results



Note. Classification accuracies in percent for each time step for RT-matched comparisons between the parts with small temporal distance within conditions (part 1.1 vs part 1.2 and part 2.1 vs part 2.2), with small temporal distance between conditions (part 1.2 vs part 2.1) and with the largest temporal distance (part 1.1 vs part 2.2). Grey areas indicate time windows in which the classification accuracies (black lines) differ significantly from the empirical chance level (grey lines).

The results show that the classifier obtains similar decoding accuracies for the partwise contrasts when the trials are RT-matched. As for the non-RT-matched data, the ANOVA for RT- matched data shows significant variations in the aggregated accuracy scores with decoding group, $F(2.24, 71.63) = 48.03$, $p < .001$, $\varepsilon = .45$, $\eta_p^2 = .60$. The accuracy scores were higher for the cross- condition decoding group with small temporal distance (part 1.2 vs 2.1, $M = 9.7\%$, $SE = 1.1\%$) compared to both within-condition decoding groups with small temporal distances (part 1.1 vs 1.2, 5.3 ± 0.6%; and part 2.1 vs part 2.2, 6.2 ± 1.0%), $p < .001$ and $p = .014$, respectively. The within- condition decoding groups did not differ significantly, $p = .954$. The scores were highest for the decoding group contrasting the temporally most distant parts (part 1.1 vs part 2.2, 18.1 ± 1.2%), followed by the decoding groups contrasting parts of medium temporal distance (part 1.1 vs 2.1, 14.1 ± 1.3% and part 1.2 vs part 2.2, 14.6 ± 1.1), and then the decoding groups comparing parts of the smallest temporal distances, all *p* values < .014. The two decoding groups that compare parts of medium temporal distance did not differ, *p* = .997. These results suggest that the behavioural differences in RT do not distort the partwise MVPA results.

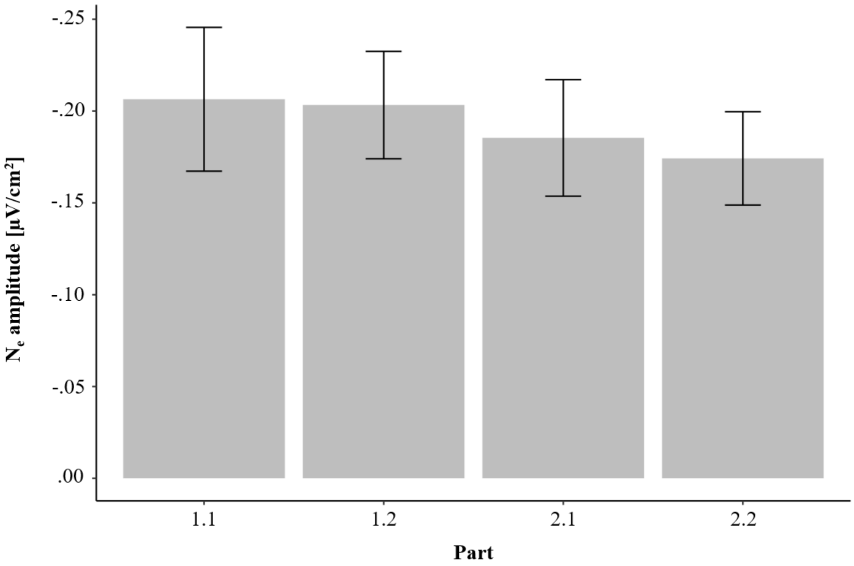
Effects of Time on Task and Counterbalancing

As the no-self-evaluation condition always preceded the self-evaluation condition, we cannot fully disentangle time on task effects from effects of self-evaluation. Unfortunately, the typical method to address time on task effects, counterbalancing, would not have resolved this problem. In a counterbalanced design the self-evaluation condition would precede the no-self-evaluation condition for half of the participants. In this order, for the preceding self-evaluation condition we could find a smaller N_e amplitude, a larger N_e amplitude or no difference in N_e amplitude between the two conditions. At first glance, a smaller N_e amplitude in the preceding self-evaluation condition (i) would be in line with our not counterbalanced results and thus suggest that in the self-evaluation condition less resources are available for fast error monitoring as effects of habituation and fatigue are controlled. However, in this case an N_e increase from the self-evaluation condition to the no-self-evaluation condition could also be explained by a (functional) carry-over effect from the self-evaluation condition to the no-self-evaluation condition (after several hundred times of self-evaluation it seems hard to not evaluate the performance even when one is not explicitly instructed to do so), by practice (which could lead to larger N_e amplitudes due to a more refined error monitoring process), and by a lowered error expectancy (which leads to larger N_e amplitudes according to the reinforcement learning theory; Holroyd & Coles, 2002). A larger N_e amplitude in the preceding self-evaluation condition (ii) would support the error significance account, but the N_e decrease towards the no-self-evaluation condition could also be an effect of habituation and fatigue. When no difference in N_e amplitude between the two conditions would be found (iii), effects of error significance, withdrawal of attentional resources, practice, habituation, and fatigue might all be at play and cancel each other out. These scenarios underline that counterbalancing cannot help to disentangle time on task effects from the effect of interest in our study.

To investigate the effect of time on task on the $N_{e/c}$ amplitude, we split the two self-evaluation conditions into two parts and conducted partwise ERP analyses. As depicted in Figure S1-6 we can observe an N_e decrease from the no-self-evaluation condition to the self-evaluation condition rather than a continuous decline in N_e amplitude across the entire experiment.

Figure S1-6

Mean and standard error of the N_e amplitude for each part



Note. N_e amplitude (mean and standard error) for the four parts of the experiment for n = 29 participants. Data sets of four participants had to be excluded from the analyses as they contained less than six error trials per part after the artefact rejection.

The ANOVA with repeated measures for the factors accuracy (error vs correct) and part (1.1, 1.2, 2.1, 2.2) only shows a significant main effect of accuracy on the $N_{e/c}$

amplitude, $F(1, 28) = 18.64$, $p < .001$, $\eta_p^2 = .40$, whereas part, $F(2.45, 68.47) = 0.81$, $p = .472$, $\varepsilon = .82$, $\eta_p^2 = .03$, and the interaction of accuracy and part, $F(3, 84) = 1.96$, $p = .126$, $\varepsilon = .89$, $\eta_p^2 = .07$, did not show significant effects. Together, these findings suggest that the tendency for an N_e decrease from the no-self-evaluation condition to the self-evaluation condition captures an effect of self-evaluation rather than a mere time on task effect.

Further Information on Study 1

The authors are grateful to Katharina Weber for supporting data collection. Open Access funding enabled and organized by Projekt DEAL. This research was supported by the German Research Foundation (STA 1035/7-1) and by Cusanuswerk e.V. (scholarship provider). The authors declare no conflict of interests. Approval was obtained from the ethics committee of the German Psychological Association. The procedures used in this study adhere to the tenets of the Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study. Participants signed informed consent regarding publishing their data.

Supplementary Materials of Study 2

P3 Mean Activity

For the P3 mean activity (motor inhibition success: 1.2 ± 0.7 µV; motor inhibition errors: 1.8 ± 0.7 µV; interference suppression success: 2.1 ± 0.6 µV; interference suppression errors: 2.0 ± 0.7 µV), there was no significant main effect of accuracy, $F(1, 29) = 0.50$, $p = .484$, $\eta_p^2 = .02$, no significant main effect of inhibition type, $F(1, 29) = 2.53$, $p = .123$, $\eta_p^2 = .08$, and no significant interaction of accuracy and inhibition type, $F(1, 29) = 2.98$, $p = .100$, $\eta_p^2 = .09$.

$P_{e/c}$ Mean Activity

For the $P_{e/c}$ mean activity, there was a significant main effect of accuracy, $F(1, 29) = 12.21$, $p = .002$, $\eta_p^2 = .30$, and inhibition type, $F(1, 29) = 8.40$, $p = .007$, $\eta_p^2 = .22$, and a tendency for an interaction of accuracy and inhibition type, $F(1, 29) = 3.94$, $p = .057$, $\eta_p^2 = .12$. Although this effect does not reach a 5% significance level, we cannot rule out that it was driven by a larger $P_{e/c}$ mean activity for motor inhibition errors (0.14 ± 0.03 µV/cm^2) than for the other three response types (motor inhibition success: 0.04 ± 0.02 µV/cm^2; interference suppression success: 0.03 ± 0.02 µV/cm^2; interference suppression errors: 0.08 ± 0.02 µV/cm^2), and a larger mean activity for interference suppression errors than for interference suppression success.

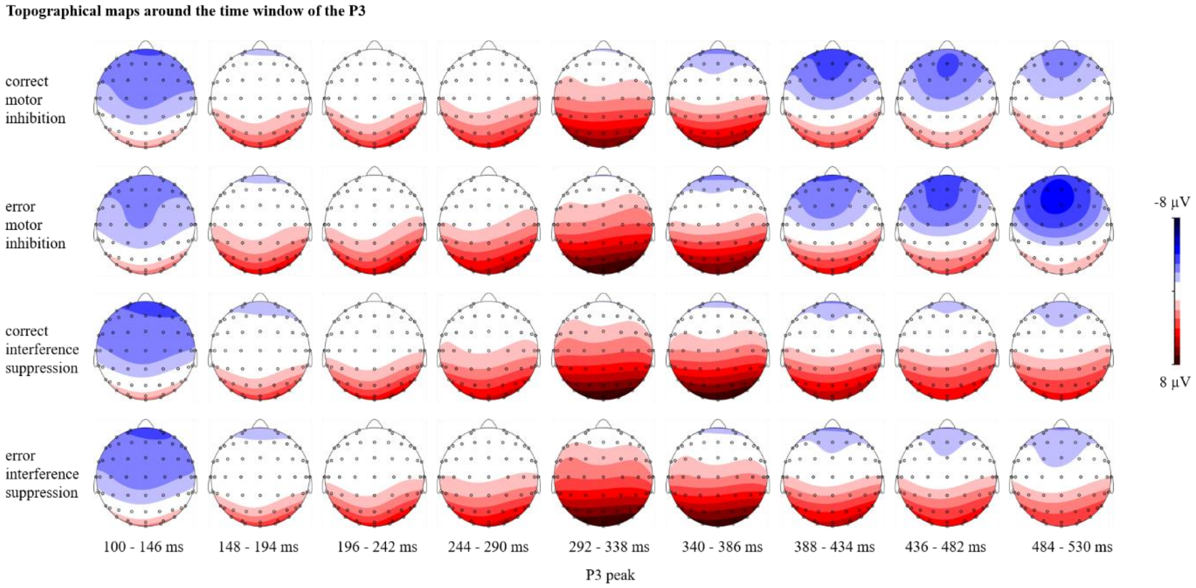
P3 Mean Amplitude at Pz

For the P3 mean amplitude measured at the Pz electrode site, there was a significant main effect of accuracy, $F(1, 29) = 7.69$, $p = .010$, $\eta_p^2 = .21$. There was no significant main effect of inhibition type, $F(1, 29) = 0.54$, $p = .468$, $\eta_p^2 = .02$, but there was a significant interaction of accuracy and inhibition type, $F(1, 29) = 10.53$, $p = .003$, $\eta_p^2 = .27$. Post hoc tests revealed that the P3 amplitude for motor inhibition success (7.0 ± 1.0 µV) was lower than for motor inhibition errors (9.1 ± 1.1 µV; *p* = .011), lower than for interference suppression success (8.2 ± 0.8 µV; *p* = .028), as well as a non-significant tendency for interference suppression errors, (8.5 ± 0.8 µV; *p* = .080). The other response types did not differ significantly, all *p* values > .524.

Topographies of the P3

Figure S2-1

Topographical maps around the time window of the P3



Paired Differences of the ERP Waveforms

Figure S2-2

Paired differences of the stimulus-locked ERP waveforms P2, N2 and P3

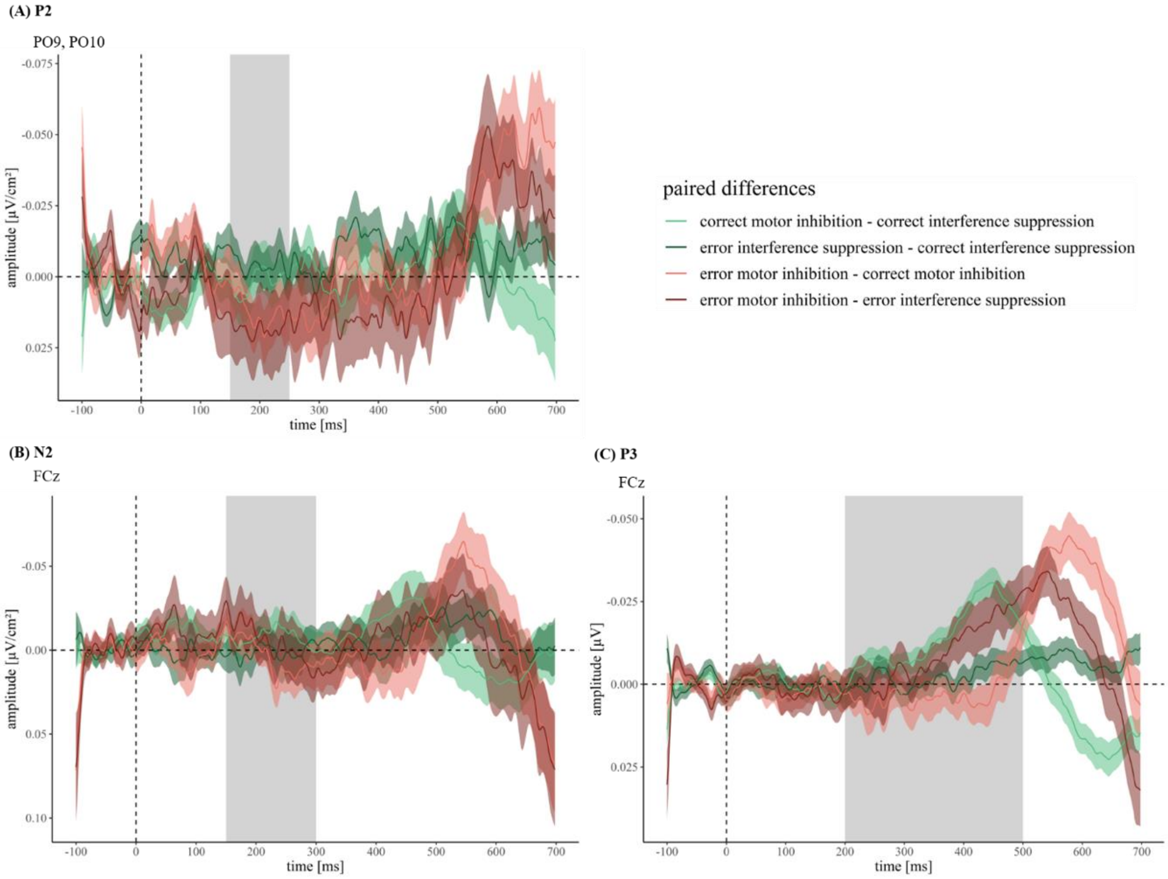
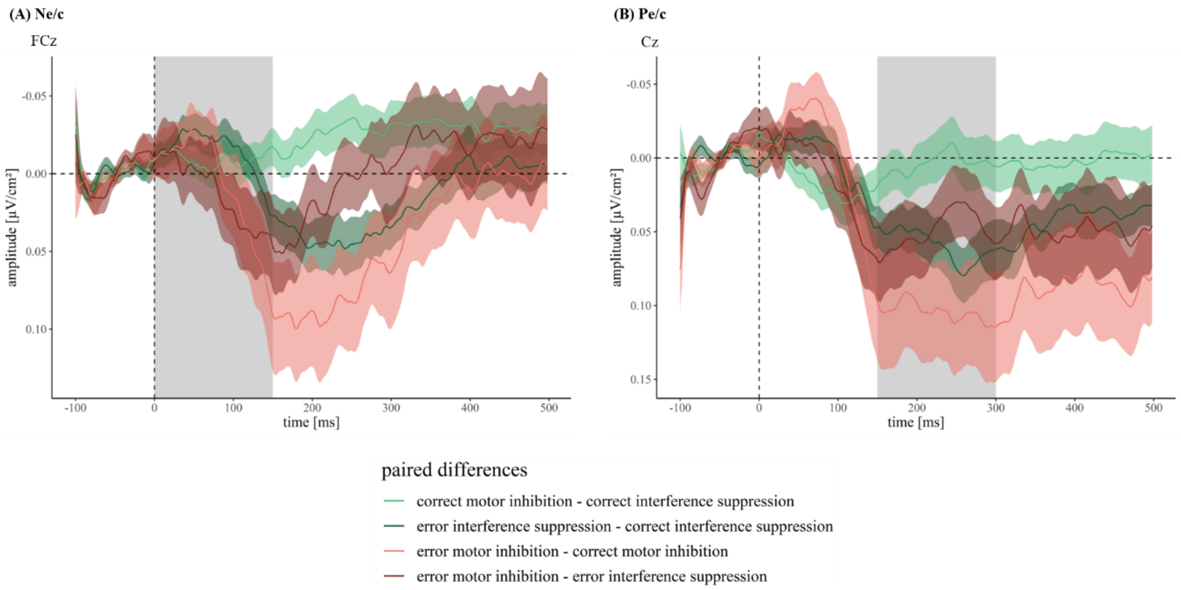


Figure S2-3

Paired differences of the response-locked ERP waveforms $N_{e/c}$ *and* $P_{e/c}$

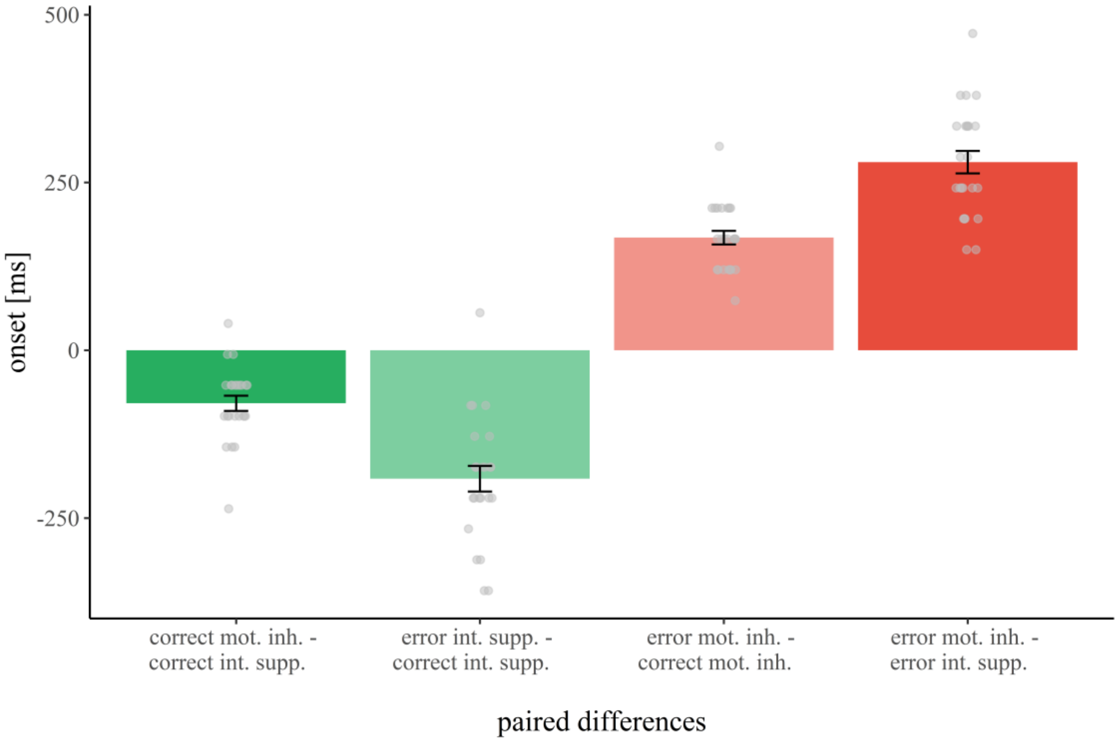


Paired Differences of the LRP onsets

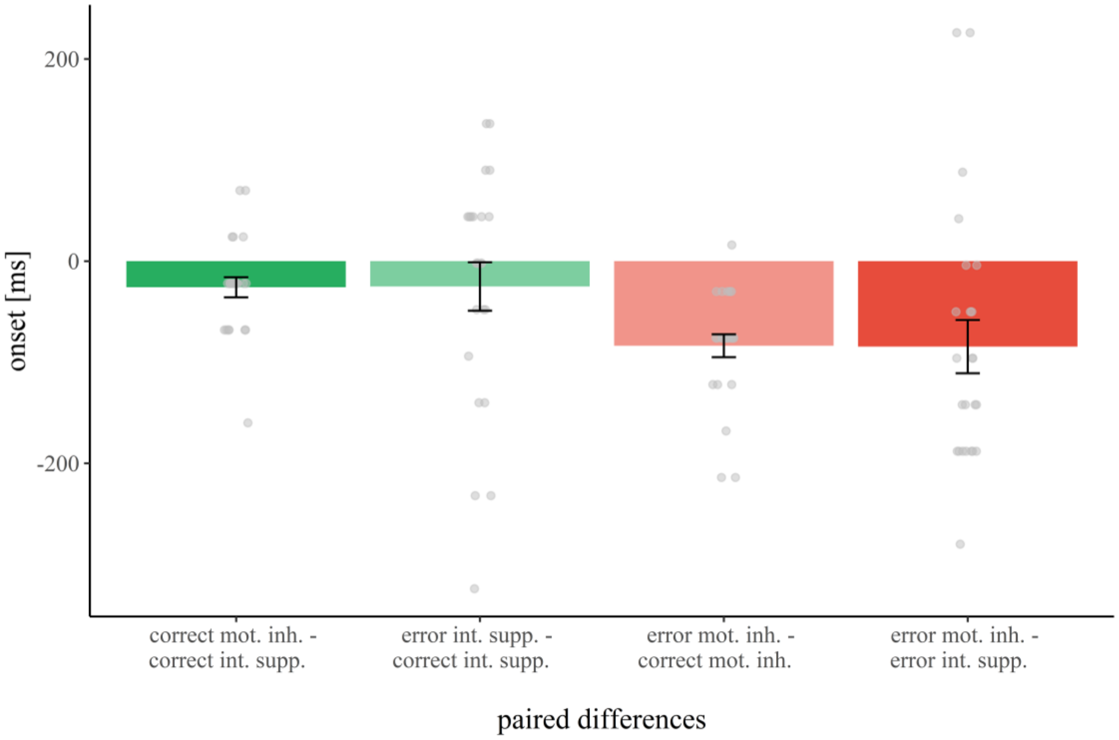
Figure S2-4

Paired differences of the stimulus-locked and response-locked LRP onsets

(A) paired differences for stimulus-locked LRP onsets



(B) paired differences for response-locked LRP onsets



Further Information on Study 2

This research was supported by the German Research Foundation (STA 1035/7-1) and by Cusanuswerk e.V. (scholarship provider). The authors declare no conflict of interests. Approval was obtained from the ethics committee of the Faculty of Human Sciences at the University of Cologne. The procedures used in this study adhere to the tenets of the Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study. Participants signed informed consent regarding publishing their data.